



DIVISION OF ELEMENTARY
& SECONDARY EDUCATION

ACT[®] **Aspire**[®]

Arkansas 2021 Technical Report

November 16, 2021

Table of Contents

Chapter 1: Introduction.....	4
1.1 Purpose and Intended Audience.....	4
1.2 Additional ACT Aspire Documentation.....	5
1.3 Proposed Uses of ACT Aspire Test Scores	8
Chapter 2: Assessment Participation	9
2.1 Spring 2021 Student Participation	9
2.2 Spring 2021 Test Administrations.....	12
2.2.1 Test mode and test accommodations	12
2.2.2 Test dates	15
2.3 Participation of Arkansas Educators in Test Development.....	16
2.3.1 April 2018 Content Review	16
2.3.2 Summer 2019 Interim Item and Form Review	17
2.3.3 Fall 2019 Development of Classroom Quizzes.....	18
Chapter 3: Achievement Summary and Trends	20
3.1 Spring 2021 Achievement Summary.....	20
3.1.1 Comparison of Mean Scores to 2019 National Norms	20
3.1.2 Comparison of ACT Readiness Benchmarks and Levels to 2019 National Norms.....	24
3.2 Achievement Trends From 2016 to 2021	30
3.3 Understanding the Impact of the COVID-19 Pandemic: Comparing Performance from 2019 and 2021	34
3.3.1 Data and Methods.....	34
3.3.2 Total Group Results	37
3.3.3 Results by Student Group.....	41
3.3.4 Summary of COVID-19 Impact on Performance	43
3.4 Differences in Performance by 2020–2021 Instructional Option: Comparing Onsite, Hybrid, and Virtual Modes of Learning.....	44
3.4.1 Data and Methods.....	44
3.4.2 Total Group Results	48
3.4.3 Results by Student Group.....	53
3.4.4 Summary of Performance Differences by Mode of Learning.....	55
Chapter 4: Technical Characteristics of the Tests	57
4.1 Test Equating.....	57

4.1.1 Equating Data Collection Design	57
4.1.2 Equating Methodology	58
4.2 Reliability and Measurement Error	58
4.2.1 Raw Score Reliability and SEM for Subject Tests	59
4.2.2 Scale Score Reliability and SEM for Subject Tests, Composite, ELA, and STEM Scores	64
4.2.3 Raw Score Reliability and SEM for Reporting Category Scores	66
4.2.4 Classification Consistency for the ACT Readiness Benchmarks and the ACT Readiness Levels	74
Chapter 5: Validity Evidence	77
5.1 Content-Oriented Evidence	77
5.1.1 Alignment Study	78
5.2 Relationships With Conceptually Related Constructs	80
5.2.1 Correlations of ACT Aspire and PARCC Scores	80
5.2.2 Correlations of ACT Aspire and 11th Grade ACT Test Scores	83
5.3 Relationships With Criteria.....	85
5.3.1 Prediction of High School Course Grades	85
5.3.2 Prediction of High School GPA and Academic Rigor	90
5.3.3 Prediction of Success on AP Exams	92
5.3.4 Prediction of First-Year College Outcomes	96
5.4 Differential Item Functioning.....	99
5.5 Depth of Knowledge Analysis	103
Chapter 6: Growth Summary.....	105
6.1 Comparison of Mean Growth Scores to National Growth Norms	105
6.2 ACT Readiness Level Transitions.....	109
6.3 Relating ACT Aspire Interim Scores to Summative SGPs	109
References.....	112
Appendix	114

Chapter 1: Introduction

1.1 Purpose and Intended Audience

ACT Aspire® Summative Assessments have been administered to Arkansas' students in 3rd through 10th grade each spring since 2016 (except for spring 2020 when testing was cancelled due to the COVID-19 pandemic). In this report, we provide information documenting the technical quality of the tests, results of testing in Arkansas, and evidence supporting intended uses of ACT Aspire test scores. The report focuses on the spring 2021 test administration, but also includes analyses of data from prior years, as well as analyses of achievement trends and student growth across years. The report is scheduled to be updated each year to reflect data from the most recent test administration and additional research pertaining to the use of ACT Aspire in Arkansas.

The intended audience for this report is individuals seeking information about the technical quality of Arkansas' ACT Aspire state assessment program. This could include educational professionals at all levels, state policymakers, and the general public. Some of the information presented in the report is of a technical nature geared towards individuals with training or experience in educational measurement or statistics.

Please note: In order to ensure the scores reported in the ACT Aspire Arkansas Technical Report (Report) can be compared to ACT Aspire national norms across multiple years, the calculations in the Report do not incorporate special scoring rules used by Arkansas for state and federal accountability such as removal of the highly-mobile population or the use of the lowest possible scale score when students have invalid scale scores or unscorable writing attempts. For the Report, students are assigned to a single race/ethnicity category. For example, for the Report, a student is only assigned to "African American" if the data indicate they are African American, and no other race, and not Hispanic. The aggregate scores in the Report may differ slightly from aggregate scores reported in <https://myschoolinfo.arkansas.gov/>.

1.2 Additional ACT Aspire Documentation

[The ACT Aspire Summative Technical Manual](#) (ACT, 2020a, hereafter referred to as *the technical manual*) presents information from a national point of view, and it contains detailed information about the ACT Aspire Summative Assessments, describes various content and psychometric aspects of the assessments, and documents a collection of evidence supporting interpretations of ACT Aspire test scores. The information contained in this report is intended to supplement the information in the technical manual, with a special focus on Arkansas-specific evidence. The technical manual and this report can be used together to assess the appropriateness of using ACT Aspire test scores for different purposes in Arkansas. This report does not duplicate content from the technical manual, unless the content is based on Arkansas-specific evidence.

Topics that are covered in the technical manual but not this technical report include:

- Test development procedures
- Test specifications
- Content standards and performance level descriptors, including:
 - ACT College and Career Readiness Standards
 - ACT Aspire Grade Level Targets for English, reading, writing, and mathematics
 - ACT Aspire Performance Level Descriptors
- Scoring procedures
- Accessibility support system and accommodations
- Test administration procedures
- Test and information security
- Interpretation of scores, readiness benchmarks, and progress indicators
- Scaling procedures

The ACT Aspire Summative Technical Manual and other forms of test documentation are available [here](#). The documentation includes FAQs, user guides, accessibility and accommodations guides, item exemplars, training resources, and guides for technology requirements and system set-up.

To help readers navigate the technical manual and this Arkansas-specific technical report, we next describe how each chapter of this report relates to content in the technical manual. The technical manual has 14 chapters and this technical report has six chapters. In Appendix [Table A.1](#), we provide a brief description of each chapter of the technical manual. The six chapters of this technical report relate to content in the technical manual as follows.

Chapter 1: Introduction. This chapter describes the purpose, scope, and organization of the technical report; and lists proposed uses of ACT Aspire test scores in Arkansas. Chapter 1 answers questions such as:

- What is the purpose and intended audience of the technical report?
- What topics are covered in the technical report?
- What topics are covered in the ACT Aspire technical manual?
- What are the proposed uses of ACT Aspire test scores?

Proposed uses and interpretations of ACT Aspire test scores are also discussed in Chapters 1 and 12 of the technical manual.

Chapter 2: Assessment Participation. This chapter documents characteristics of Arkansas students who participated in testing, summarizes information on test administrations (including use of test forms with accommodations), and provides information on the participation of Arkansas educators in item writing and item and form reviews. Chapter 2 answers questions such as:

- How many students took ACT Aspire tests in spring 2021?
- How many students from each student group participated in testing?
- What types of testing accommodations were utilized?
- What were the most popular test dates?
- How have educators from Arkansas been engaged in test development activities?

Information on frequency of use of test forms with accommodations is also provided in Chapter 5 of the technical manual.

Chapter 3: Achievement Summary and Trends. This chapter presents summary statistics on ACT Aspire scores and readiness levels for the 2021 ACT Aspire administration in Arkansas, with comparisons to national norms and to prior years. It answers questions such as:

- What were the average test scores in spring 2021?
- What percent of Arkansas students are on target for college and career readiness?
- How do Arkansas' scores from 2021 compare to national averages from 2019?
- Have scores improved since the beginning of the assessment program in 2016?

- To what extent did scores decline during the COVID-19 pandemic?
- How were different instructional options (Onsite, Hybrid, Virtual) related to performance?

Estimates of national norms for ACT Aspire scores are presented in Chapter 8 of the technical manual.

Chapter 4: Technical Characteristics of the Tests. This chapter presents information on procedures for equating tests. It also presents estimates of reliability, standard error of measurement, and classification consistency for the 2021 ACT Aspire administration in Arkansas. It answers questions such as:

- What procedures are used to ensure that ACT Aspire Summative test scores are comparable across different years and test forms?
- How reliable are ACT Aspire Summative test scores?
- Are the test scores reliable for different groups of students?
- To what extent are students classified consistently with respect to being on target for college and career readiness?

Similar information for all ACT Aspire summative users is provided in Chapters 10 and 11 of the technical manual.

Chapter 5: Validity Evidence. This chapter summarizes an independent study of the alignment of ACT Aspire to Arkansas' state standards. It also presents Arkansas-specific criterion-related validity evidence, differential item functioning (DIF) analysis, and analyses related to depth of knowledge (DOK) levels. Chapter 5 answers questions such as:

- How well is ACT Aspire test content aligned to the state's academic standards?
- What is the relationship between ACT Aspire scores and performance in high school courses?
- How well do ACT Aspire scores predict ACT test scores?
- Are ACT Aspire scores predictive of first-year college outcomes?
- Do ACT Aspire test items function similarly for different groups of examinees?

The independent alignment study is not presented in the technical manual, but Chapter 1 of the technical manual includes some general discussion of alignment. Arkansas-specific criterion-related validity evidence is presented in both this technical report and Chapter 12 of the technical manual. A national DIF analysis (not specific to Arkansas examinees) is presented in Chapter 13 of the technical manual.

Chapter 6: Growth Summary. This chapter summarizes Arkansas-specific growth data, with comparisons to national norms. It answers questions such as:

- How does ACT Aspire support interpretations of academic growth?
- How much do student's scores typically increase in one year?
- How do Arkansas' growth scores compare to national averages?
- How do ACT Aspire Interim test scores relate to growth in summative scores?

Chapter 14 of the technical manual presents national data summarizing gain scores and Student Growth Percentiles.

1.3 Proposed Uses of ACT Aspire Test Scores

One purpose of this report is to provide evidence supporting proposed uses of ACT Aspire test scores in Arkansas, which include:

- To measure progress toward meeting the Arkansas academic standards for grades 3-8 and high school in English Language Arts (English, reading, and writing), mathematics, and science
- To determine if Arkansas students are on target for college and career readiness
- To assess how well Arkansas schools and districts are preparing students for college and careers by meeting grade level standards (school and district accountability)
- To inform students' readiness for advanced high school coursework
- To understand student and group performance relative to national norms

For example, the use of ACT Aspire scores for accountability is supported by content evidence, studies examining alignment of ACT Aspire with the state's academic standards, evidence from standard setting (including development of the ACT Readiness Benchmarks), and additional evidence presented in this technical report or the technical manual. Test users may develop additional uses that are not listed here and may need to collect additional evidence to support them.

ACT Aspire includes Summative, Periodic, and Classroom Assessments. While all three types of assessments have been used in Arkansas, this report focuses on the Summative Assessments.

Chapter 2: Assessment Participation

ACT Aspire Summative Assessments are intended for students in 3rd through 10th grade and are designed to measure key college and career readiness constructs in a way that recognizes that skills are not isolated to specific grades, but rather should progress across grades. Assessments are available for five domains (subjects): English, mathematics, reading, science, and writing. Separate assessments are used for each grade and subject, with the exception of 9th and 10th grade, which are both assessed using test forms developed for early high school.

In this chapter we document characteristics of Arkansas students who participated in testing in spring 2021 and provide information on test administrations, including mode of testing, accommodations, and dates of testing. Because this report does not use the special scoring rules used by Arkansas for state and federal accountability, the number of students reported may be slightly different than those reported for state and federal accountability. We also provide information on the participation of Arkansas educators in ACT Aspire test development activities.

2.1 Spring 2021 Student Participation

Table 2.1 provides the number of students for whom scores were reported in spring 2021. Overall, 286,761 students had at least one score reported. In addition to the total number of students, counts are also provided for each subject area. Among students with at least one reported score, the overwhelming majority—96.9%—had scores reported for all five subject areas. The subject with the highest rate of missing scores was writing, with 2.7% of all students missing a writing score.

Table 2.1. Number of Students Tested in Spring 2021, by Grade Level and Subject

Grade level	Total	Subject				
		English	Mathematics	Reading	Science	Writing
3	34,845	34,800	34,813	34,818	34,794	32,658
4	34,832	34,793	34,805	34,796	34,799	33,138
5	35,431	35,392	35,395	35,403	35,381	34,649
6	36,591	36,500	36,499	36,515	36,479	35,824
7	37,592	37,498	37,469	37,494	37,469	36,966
8	37,789	37,647	37,671	37,680	37,643	37,375
9	35,456	35,268	35,267	35,255	35,278	34,841
10	34,225	33,966	33,970	33,997	33,944	33,535
Total	286,761	285,864	285,889	285,958	285,787	278,986

Note. The counts do not incorporate special scoring rules used by Arkansas for state and federal accountability.

Evidence of the validity and fairness of using ACT Aspire assessments for its intended purposes for different groups of students is provided in later chapters of this report and the technical manual. The groups that are included for accountability reporting include the following major racial and ethnic student groups and educationally at-risk student groups: African American, Hispanic, White, Economically Disadvantaged, Students with Disabilities, and English Learners. Note that group membership is determined from data collected by the ACT Aspire registration system and may differ from other state records. Also, the Students with Disabilities group includes students with an individualized education plan (IEP), Section 504 plan, or other accommodations plan.

Students are assigned to a race/ethnicity category based on the following rules, applied sequentially:

- 1) If of Hispanic ethnicity, the student is categorized as *Hispanic*.
- 2) Otherwise, if a student is assigned “Yes” to two or more race indicators, they are categorized as *Two or more races*.
- 3) Otherwise, if a student is assigned “Yes” to one race indicator, they are categorized as that race (*African American, Asian, Native American, Native Hawaiian/Other Pacific Islander, or White*).
- 4) Otherwise, the student is categorized as *Missing race/ethnicity*.

In Table 2.2, we provide the number of students in each group, by grade level. In addition to the major groups identified for accountability reporting, we provide counts for additional groups, including gender and smaller racial/ethnic groups, and Migrant. Table 2.3 provides the percentage of all students belonging to each group.

Table 2.2. Number of Students Tested in Spring 2021, by Student Group and Grade Level

Group	Grade level							
	3	4	5	6	7	8	9	10
Gender								
Female	16,808	17,062	17,295	17,768	18,394	18,562	17,318	16,823
Male	18,037	17,770	18,136	18,823	19,198	19,227	18,117	17,354
Missing gender	0	0	0	0	0	0	21	48
Race/ethnicity								
African American	6,752	6,782	7,009	7,152	7,507	7,291	6,583	6,323
Asian	651	586	579	662	663	595	548	641
Hispanic	4,794	5,041	5,032	5,371	5,315	5,437	5,027	4,616
Native American	213	185	212	225	204	219	221	220
Native Hawaiian/OPI	409	402	384	365	340	356	332	292
Two or more races	1,317	1,240	1,179	1,158	1,268	1,233	1,105	999
White	20,705	20,594	21,034	21,657	22,293	22,657	21,618	21,081
Missing race/ethnicity	4	2	2	1	2	1	22	53
Migrant	274	270	259	280	276	239	203	193
Econ. Disadvantaged	23,190	23,027	23,378	23,678	23,914	23,625	21,143	19,697
Students with Disabilities	7,493	7,611	7,780	7,933	7,685	7,316	6,811	6,247
English Language Learner	3,192	2,934	2,589	2,552	2,300	2,399	2,226	2,235

Note. Bold indicates accountability reporting group. OPI = Other Pacific Islander; Econ. = Economically. The counts do not incorporate special scoring rules used by Arkansas for state and federal accountability.

Table 2.3. Percentage in Student Groups, by Grade Level

Group	Grade level							
	3	4	5	6	7	8	9	10
Gender								
Female	48.2	49.0	48.8	48.6	48.9	49.1	48.8	49.2
Male	51.8	51.0	51.2	51.4	51.1	50.9	51.1	50.7
Missing gender	0	0	0	0	0	0	0.1	0.1
Race/ethnicity								
African American	19.4	19.5	19.8	19.5	20.0	19.3	18.6	18.5
Asian	1.9	1.7	1.6	1.8	1.8	1.6	1.5	1.9
Hispanic	13.8	14.5	14.2	14.7	14.1	14.4	14.2	13.5
Native American	0.6	0.5	0.6	0.6	0.5	0.6	0.6	0.6
Native Hawaiian/OPI	1.2	1.2	1.1	1.0	0.9	0.9	0.9	0.9
Two or more races	3.8	3.6	3.3	3.2	3.4	3.3	3.1	2.9
White	59.4	59.1	59.4	59.2	59.3	60.0	61.0	61.6
Missing race/ethnicity	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2
Migrant	0.8	0.8	0.7	0.8	0.7	0.6	0.6	0.6
Econ. Disadvantaged	66.6	66.1	66.0	64.7	63.6	62.5	59.6	57.6
Students with Disabilities	21.5	21.9	22.0	21.7	20.4	19.4	19.2	18.3
English Language Learner	9.2	8.4	7.3	7.0	6.1	6.3	6.3	6.5

Note. Bold indicates accountability reporting group. OPI = Other Pacific Islander; Econ. = Economically. The percentages do not incorporate special scoring rules used by Arkansas for state and federal accountability.

2.2 Spring 2021 Test Administrations

2.2.1 Test mode and test accommodations

A variety of accessibility supports, tools, and options are available to ensure that ACT Aspire assessments are administered in accessible and standardized ways. Multiple levels of accessibility are available, including default embedded tools, open-access tools, and full accommodations to allow students with disabilities to participate in testing. Tools and accommodations are available for both online and paper tests. Some, but not all, tools or accommodations require a different type of test form.

For more information about accessibility supports, tools, and options for ACT Aspire Summative testing, please see [ACT Aspire Accessibility Supports Guide](#).

Tests were primarily administered in an online mode: Across all subject areas and grade levels, only 996 tests out of 1,422,494 (0.07%) were administered using paper.

Table 2.4 reports the frequency of each type of accommodation provided, by subject area. The table also provides the number of students receiving at least one accommodation, and the number of students who tested without accommodations. The most common types of accommodations included special seating or grouping (n = 202,620 tests), extra time (n = 157,552 tests), and English text-to-speech audio (n = 143,899 tests). Note that testing with text-to-speech English audio supports requires a different test form, and Arkansas policy does not allow this accommodation for the English and reading tests.

Table 2.4. Accommodations, by Subject

Accommodation Type	English	Math	Reading	Science	Writing
Total	285,864	285,889	285,958	285,787	278,986
None	223,576	221,753	223,728	221,648	219,685
Accommodated	62,288	64,136	62,230	64,139	59,301
Abacus	0	33	0	0	0
American Sign Language Directions Only	*	*	*	*	*
American Sign Language Full Translation	0	60	0	60	39
Audio Environment	71	70	70	71	58
Braille Contracted American Edition EBAE	*	*	*	*	*
Braille Contracted Unified English UEB	18	22	18	22	22
Breaks	*	*	*	*	*
Breaks Supervised Each Day	1,365	1,371	1,371	1,372	1,183
Cued Speech	0	*	0	*	*
Custom Masking	*	*	*	*	*
Dictate Responses	100	100	99	101	84
Electronic Spell Checker	0	*	0	*	*
English TTS Audio	0	49,438	0	49,447	45,014
English TTS Audio Orienting Description	0	*	0	*	*
Extra Time 1.5	27,100	7,813	27,017	7,797	7,542
Extra Time 2.0	3,083	1,028	3,080	1,029	1,012
Extra Time 2.5	*	*	*	*	*
Extra Time 3.0	28,399	4,692	28,454	4,681	4,598
Extra Time 4.0	73	17	73	23	23
Home Administration	*	*	*	*	*
Human Reader English	0	100	0	99	74
Human Reader English Orienting Desc.	0	13	0	13	13
Individual Administration	826	825	824	828	609
Keyboard AAC Local Print	21	21	21	21	20
Large Print	31	31	31	31	28
Location for movement	96	95	96	96	77
Other Setting	1,024	1,016	1,016	1,010	930
Physical Motor Equipment	14	14	14	14	12
Respond in Test Booklet / Separate Paper	28	27	27	27	27
Signed Exact English Full Translation	0	*	0	*	*
Spanish Text Audio	0	*	0	*	*
Special Seating or Grouping	8,139	8,162	8,152	8,160	7,833
Translated Test Directions Only	254	255	254	257	170
Visual Environment	12	12	12	12	12
Word 2 Word Dictionary	*	709	*	708	581

Note. * Indicates that count is greater than 0 and less than 10.

2.2.2 Test dates

Figure 2.1 provides the number of ACT Aspire tests taken by date. The tests were taken from April 5th to June 2nd, with the vast majority taken by May 10th. Most tests were taken the week of April 19–23, followed by the week of April 12–16, the week of April 5–9, and the week of April 26–30.

Among students who took all five tests, a small percentage (2.1%) took all tests on the same day. Most students tested over the course of two (32.9%), three (31.1%), four (18.7%), or five days (15.1%).

The order that tests were administered varied, with reading, English, and writing usually administered before mathematics and science. Among all reading tests, 52.1% were taken the first day of testing for a student, compared to 49.2% for English, 36.6% for writing, 32.5% for mathematics, and 22.1% for science. Conversely, among all science tests, 54.8% were taken on the last day of testing for a student, compared to 41.4% for writing, 34.3% for English, 26.4% for mathematics, and 19.9% for reading.

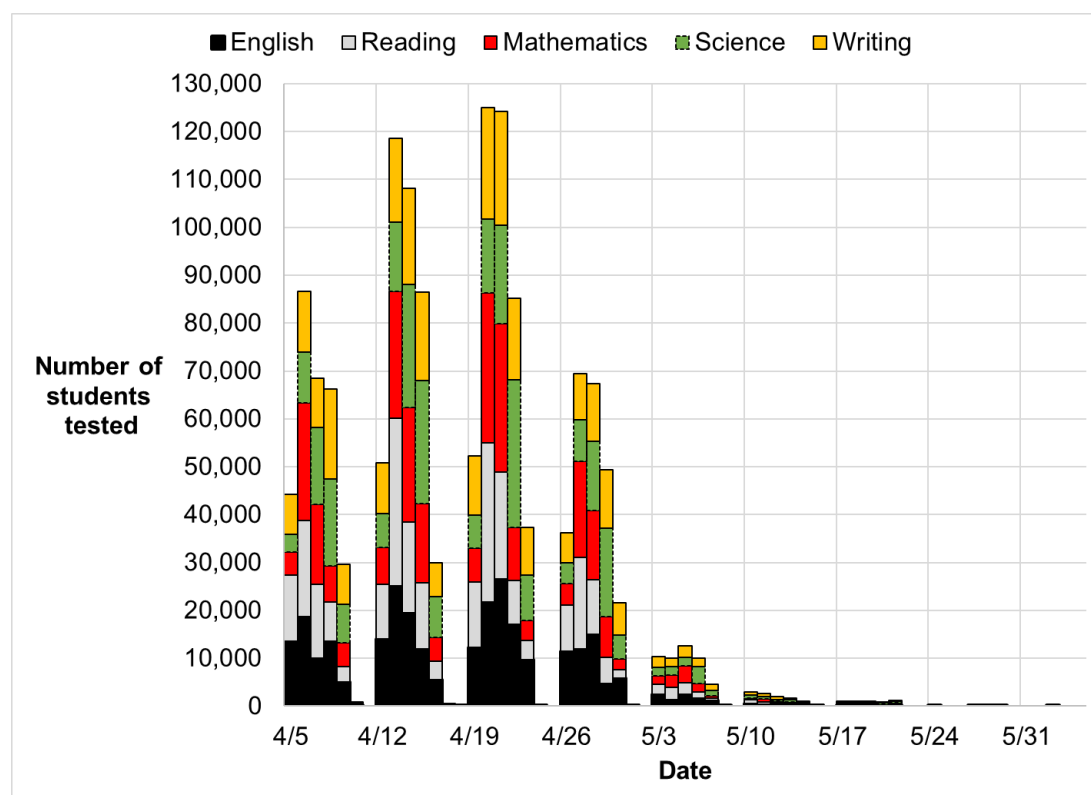


Figure 2.1. Number of students tested, by date and subject

2.3 Participation of Arkansas Educators in Test Development

ACT recruits item writers to form a diverse group of individuals with extensive content and pedagogical knowledge who have experience teaching grades 3–10. Item writers include teachers, subject specialists, instructional coaches, curriculum coordinators, and department chairs. All items undergo rigorous content reviews to ensure that they elicit sufficient student evidence, are developmentally appropriate, and that their contents and contexts are error-free. External experts participate in fairness reviews to ensure that items and tasks are not biased.

Teachers and other educational professionals from Arkansas have served as item writers and reviewers of ACT Aspire items and test forms. Next, we describe three recent test development events that included educators from Arkansas.

2.3.1 April 2018 Content Review

In April 2018, a two-day item content review session was held at the ACT campus in Iowa City, Iowa. Among the participants were 15 individuals from Arkansas, recruited by the Arkansas State Department of Education. Characteristics of participants from Arkansas are included in Table 2.5.

Table 2.5. Arkansas Participants in April 2018 Content Review

Characteristic	N
Position	
Teacher	6
Instructional Coach/Facilitator	6
Department Chair	1
Other	2
Years of teaching	
6–10	6
11–20	6
>20	3
Subject and grade band	
English, 6–10	0
Reading, 3–5	2
Reading, 6–10	1
Mathematics, 3–5	2
Mathematics, 6–10	3
Science, 3–5	4
Science, 6–10	3

Participants reviewed ACT Aspire Summative test items prior to their selection for use on operational test forms. Review activities were organized according to content area (English, reading, mathematics, and science) and grade band (3–5 and 6–10). ACT Content Specialists facilitated separate break-out sessions for each content area / grade band. Packets of items were distributed at the beginning of each session and collected at the end of each session. Each group openly discussed each item under review. Participants were expected to read each item and review for the following:

- Grade-level appropriateness in terms of information and ideas presented and skills assessed
- Content accuracy
- Answer key propriety (multiple-choice items should have one and only one best answer, also known as the key)
- Plausibility of distractors
- Bias and/or sensitivity concerns

After the workshop, the ACT Content Specialists in each content domain collated feedback from their specific educator participants, reviewed the feedback, and revised item content as deemed appropriate.

2.3.2 Summer 2019 Interim Item and Form Review

In summer 2019, educators from Arkansas and other states participated in the review of new ACT Aspire Interim items and test forms. ACT recruited educators from grades 3–9 to participate in the reviews to ensure the accuracy of test content and guarantee that each form meets the breadth and depths of the standards that it measures. Reviews were completed in ACT's web-based system known as RIMS. Directions regarding content review best practices were provided by ACT staff and the reviews were conducted remotely.

Participants were required to sign and return a consulting security agreement, complete online training, and/or carefully review guidelines and process documentation. Content reviewers were expected to read each passage and/or item and review for the following:

- Answer key propriety
- Unwanted cluing: Does content in one item provide the answer for another item in a set?
- Comprehensibility and effectiveness of wording in each item
- Grade-level appropriateness in terms of information and ideas presented and skills assessed

- Readability and accessibility of graphics, equations, diagrams, and tables
- Content reviewers did not review for the following:
 - Format, style, grammar, punctuation, and spelling
 - Fairness, bias, and sensitivity; there is a separate review specifically focused on these areas.

Characteristics of the 46 individuals from Arkansas who served as reviewers are included in Table 2.6.

After the item and form reviews were completed, three-hour panel sessions (one for each subject and grade band) were conducted via Skype with 26 of the participants from Arkansas to discuss their specific feedback with ACT content experts.

Table 2.6. Arkansas Participants in Summer 2019 Interim Form Review

Characteristic	N
Position	
Teacher	30
Instructional Coach/Facilitator	12
Specialist	2
Principal/Assistant Principal	1
Other	1
Years of teaching	
3–5	3
6–10	11
11–20	20
>20	12
Subject and grade band	
ELA, 3–5	4
ELA, 6–7	5
ELA, 8–9	7
Mathematics, 3–5	3
Mathematics, 6–7	5
Mathematics, 8–9	6
Science, 3–5	6
Science, 6–7	5
Science, 8–9	5

2.3.3 Fall 2019 Development of Classroom Quizzes

In September 2019, a workshop was conducted to develop additional quizzes for ACT Aspire Classroom. ACT recruited educators from grades 3–8 to participate in a five-day

item writing workshop at the ACT campus. During the workshop, participants representing four content domains (English, mathematics, reading, and science) worked individually and collaboratively to develop a pool of items to be used on ACT Aspire Classroom quizzes.

Participants were expected to complete training prior to the workshop, including 2–4 hours of pre-workshop item-writing training and participation in an online question and answer meeting the week before the workshop. Participants must have at least two years of teaching experience in grade bands 3–5 or 6–8 and the ability to provide concise verbal and written feedback. Characteristics of the individuals from Arkansas who registered for the workshop are included in Table 2.7.

Table 2.7. Arkansas Participants Registered for September 2019 Classroom Quiz Workshop

Characteristic	N
Position	
Teacher	25
Instructional Coach/Facilitator	13
Specialist	9
State Department of Education Staff	6
Department Chair	1
District Test Coordinator	1
Subject and grade band	
English, 3–5	8
English, 6–8	6
Reading, 3–5	7
Reading, 6–8	6
Mathematics, 3–5	8
Mathematics, 6–8	6
Science, 3–5	7
Science, 6–8	7

Chapter 3: Achievement Summary and Trends

This chapter presents summary statistics on ACT Aspire scores and readiness levels for the 2021 ACT Aspire administration in Arkansas, with comparisons to national norms and to prior years.

As discussed in [Chapter 2](#), Arkansas administers tests for all five subject areas: English, mathematics, reading, science, and writing. Scale scores are generated for each of the five tests and are reported for four of the five tests (all but writing). In addition to scale scores for four subjects, three combined scores are reported:

- The **ELA score** is the average of the English, reading, and writing scale scores.
- The **STEM score** is the average of the mathematics and science scale scores.
- The **Composite score** is the average of the English, mathematics, reading, and science scale scores.

This chapter summarizes achievement and trends using the scale scores and combined scores, and the corresponding ACT Readiness Benchmarks and ACT Readiness Levels.

As mentioned in the Introduction, in order to compare performance in Arkansas to ACT Aspire national norms across multiple years, this report does not incorporate special scoring rules used by Arkansas for state and federal accountability such as removal of the highly-mobile population or the use of the lowest possible scale score when students have invalid scale scores or unscoreable writing attempts. The special scoring rules have an impact on sample sizes, test score means and standard deviations, and the percentage of students scoring in each achievement level.

3.1 Spring 2021 Achievement Summary

3.1.1 Comparison of Mean Scores to 2019 National Norms

We begin by examining the mean scores for Arkansas from spring 2021 and comparing them to the latest national norms. This analysis addresses the question: “How does the academic achievement of Arkansas students in 2021 compare to the pre-pandemic academic achievement of students across the nation?”

ACT periodically conducts a national norming study to produce updated estimates of percentile ranks and mean scores for each reported scale score, by grade level. The norms used for this report are based on the 2019 norming study, which used national data from spring 2017, spring 2018, and spring 2019. The norming study was not done

in 2020 because ACT Aspire Summative Assessments were not administered in spring 2020 due to the COVID-19 pandemic.

The norming samples include students from both public and nonpublic schools, including those from Arkansas. ACT Aspire-tested students are not necessarily representative of the national population of students in grades 3–10. To support interpretations of nationally representative norms, weights are assigned to the samples so that they are more similar to the national population on school affiliation (public vs. non-public) and, among public schools, race/ethnicity and academic achievement. More details on the norming study methodology are provided in Chapter 8 of the technical manual.

For each subject and grade level, the mean and standard deviation of scale scores are provided in Table 3.1. Statistics are provided for the spring 2021 Arkansas administration and the 2019 national norms. The ***d*** statistic of Table 3.1 measures the difference between each Arkansas mean score and the estimated national mean score. It is calculated as the difference between the two means, divided by the estimated national standard deviation. Positive values of ***d*** indicate that the Arkansas mean score is larger than the estimated national mean score.

For each subject and grade level, the Arkansas mean was less than the pre-pandemic national mean. Relative to the pre-pandemic national norms, Arkansas scores were lowest for 5th (***d*** = -0.47) and 6th grade (***d*** = -0.45) mathematics and highest for 8th grade reading (***d*** = -0.08) and 7th grade English (***d*** = -0.09).

When the Arkansas mean scores are different than the national mean scores by 0.10 standard deviations and larger ($|d| \geq 0.10$), we consider it a substantive difference. Using this rule, the mean scores for Arkansas are substantively lower than the pre-pandemic national norms for 54 of the 56 combinations of subject/score and grade level. Arkansas' 2021 test scores were impacted by the learning disruptions caused by the COVID-19 pandemic, whereas the 2019 national norms were not. If national norms were available for 2021, we would expect to find a significant decline in national norms relative to 2019. Using Arkansas' 2019 test scores, we only found 5 cases (out of 56) for which the Arkansas mean scores were substantively lower than the national norms.

In later sections of this report, we present additional analyses of Arkansas' scores from 2019 and 2021 to better understand possible impacts of the pandemic. In addition to the pandemic, there could be many other reasons for differences in Arkansas performance across subjects and grade levels, relative to national norms. In this report, we do not attempt to explain the differences.

Table 3.1. 2021 Scale Score Summary Statistics, by Subject and Grade Level

Subject	Grade level	2021 Arkansas		2019 National Norms		<i>d</i>
		Mean	SD	Mean	SD	
English	3	415.4	6.3	417.2	6.4	-0.28
	4	419.0	6.7	420.3	6.6	-0.19
	5	421.9	6.9	423.1	7.0	-0.17
	6	423.8	8.2	425.3	8.4	-0.18
	7	426.7	8.7	427.5	8.8	-0.09
	8	426.5	9.6	427.8	9.4	-0.13
	9	427.0	10.3	428.8	10.4	-0.18
	10	428.9	10.9	431.1	10.9	-0.20
Mathematics	3	412.0	4.5	413.5	4.4	-0.36
	4	414.8	4.5	416.3	4.3	-0.36
	5	415.5	4.8	418.1	5.5	-0.47
	6	418.3	5.6	421.1	6.3	-0.45
	7	419.8	7.3	421.4	7.8	-0.21
	8	422.2	7.8	423.9	8.3	-0.20
	9	422.4	8.0	425.2	8.6	-0.32
	10	424.0	8.6	427.0	9.4	-0.32
Reading	3	411.8	5.4	413.1	5.5	-0.24
	4	414.7	6.3	415.6	6.1	-0.16
	5	416.5	6.6	417.6	6.4	-0.17
	6	418.0	6.8	419.4	6.9	-0.20
	7	419.0	6.8	420.2	6.7	-0.18
	8	422.2	7.7	422.8	7.3	-0.08
	9	421.2	7.9	422.5	7.9	-0.16
	10	422.3	8.3	423.5	8.2	-0.15

Table 3.1. (continued)

Subject	Grade level	2021 Arkansas		2019 National Norms		<i>d</i>
		Mean	SD	Mean	SD	
Science	3	414.0	6.9	415.4	6.7	-0.21
	4	416.6	6.8	418.0	6.7	-0.21
	5	418.3	6.6	420.1	6.5	-0.28
	6	419.5	7.2	421.4	7.2	-0.26
	7	421.6	7.9	422.6	8.0	-0.12
	8	423.3	8.4	424.4	8.3	-0.13
	9	424.9	8.5	425.7	8.8	-0.10
	10	426.1	9.0	427.2	9.5	-0.12
Composite	3	413.4	5.3	414.9	5.2	-0.28
	4	416.4	5.5	417.6	5.4	-0.23
	5	418.2	5.5	419.7	5.7	-0.27
	6	420.0	6.3	421.9	6.5	-0.28
	7	421.9	7.0	423.1	7.1	-0.17
	8	423.7	7.6	424.9	7.5	-0.16
	9	424.0	7.9	425.8	8.1	-0.21
	10	425.5	8.4	427.4	8.7	-0.22
ELA	3	416.1	5.2	417.7	5.1	-0.32
	4	419.6	5.5	420.3	5.2	-0.15
	5	421.0	5.6	422.3	5.9	-0.22
	6	423.4	6.5	424.5	6.4	-0.17
	7	423.4	6.6	424.4	6.6	-0.15
	8	424.8	6.9	425.6	6.8	-0.12
	9	424.9	7.4	426.3	7.5	-0.20
	10	426.2	7.7	427.7	7.8	-0.20
STEM	3	413.3	5.4	414.7	5.2	-0.28
	4	415.9	5.3	417.4	5.3	-0.27
	5	417.2	5.2	419.4	5.7	-0.39
	6	419.2	6.0	421.4	6.3	-0.36
	7	421.0	7.2	422.3	7.4	-0.18
	8	423.0	7.7	424.5	7.9	-0.18
	9	423.9	7.8	425.8	8.2	-0.23
	10	425.3	8.4	427.5	9.0	-0.24

Note. SD = standard deviation; *d* = (2021 Arkansas mean – 2019 National mean) / National SD. The statistics do not incorporate special scoring rules used by Arkansas for state and federal accountability.

3.1.2 Comparison of ACT Readiness Benchmarks and Levels to 2019 National Norms

The ACT College Readiness Benchmarks are the scores on the ACT® test associated with a 50% chance of earning a B or higher grade in common first-year credit-bearing college courses. ACT College Readiness Benchmarks have been developed for English (18), mathematics (22), reading (22), science (23), ELA (20), and STEM (26).

More information on the ACT College Readiness Benchmarks is [available here](#).

The ACT Readiness Benchmarks are the minimum ACT Aspire scores (grades 3–10) for which students are on target to meet or exceed the ACT College Readiness Benchmarks when they are in 11th grade. Thus, the ACT Readiness Benchmarks can be interpreted as the minimum scores associated with being on-target for college readiness. ACT Readiness Benchmarks are reported for each subject and grade level.

For English, mathematics, reading, science, and ELA scores, ACT Readiness Levels are used to further classify student achievement as:

- **In Need of Support** if the score is greater than two standard errors of measurement (SEM) below the ACT Readiness Benchmark.
- **Close** if the score is below the ACT Readiness Benchmark, but within two SEMs of the Benchmark.
- **Ready** if the score is equal to the ACT Readiness Benchmark, or above and within two SEMs of the Benchmark.
- **Exceeding** if the score is greater than two SEMs above the ACT Readiness Benchmark.

Similar to the analysis of mean scores, the percentage of Arkansas students scoring at each ACT Readiness Level can be compared to the pre-pandemic national norms (Table 3.2). For STEM, note that the table only provides the percentage of students who met the ACT Readiness Benchmark because ACT Readiness Levels have not been set for the STEM score. For ELA, note that ACT Readiness Levels are not reported on ACT Aspire score reports, but are used for purposes of classifying the ELA performance of Arkansas' students. To permit comparisons to the 2019 national norms, the percentages do not incorporate the special scoring rules used by Arkansas for state and federal accountability discussed earlier.

Similar to the national norms, students in Arkansas are more likely to meet the English Benchmark, and least likely to meet the STEM Benchmark. The ACT STEM Benchmark was derived using college courses most commonly taken by students in STEM-related majors; the courses included Calculus, Biology, Chemistry, and Engineering. Because STEM-related coursework tends to be more difficult, the resulting Benchmark is substantially higher than the Benchmarks for other subjects.

The “Diff. Bench” column of Table 3.2 shows the difference in Benchmark attainment rates for Arkansas students relative to the 2019 national norms, with positive values indicating that Arkansas students outperformed the national norm. Across all subjects and grade levels, Benchmark attainment for Arkansas students was lower than that of the pre-pandemic national norm. Relative to the norms, Arkansas Benchmark attainment was lowest for 6th grade mathematics (-17%) and 5th grade mathematics (-16%) and highest for 8th grade reading (-2%) and 7th grade English (-2%).

Arkansas’ Benchmark attainment rates for 2021 were impacted by the learning disruptions caused by the COVID-19 pandemic, whereas the 2019 national norms were not. If national norms were available for 2021, we would expect to find a significant decline in Benchmark attainment nationally relative to 2019.

Table 3.2. Percentage Meeting ACT Readiness Levels and Benchmarks, by Subject and Grade Level

Subject	Grade level	2021 Arkansas					2019 National Norms					Diff. Bench.
		INS	Close	Ready	Exceed.	Met Bench.	INS	Close	Ready	Exceed.	Met Bench.	
English	3	6	32	27	35	62	5	21	28	46	74	-13
	4	13	25	30	33	63	7	23	32	38	70	-7
	5	7	23	38	31	69	5	22	37	37	73	-4
	6	7	25	32	36	68	6	20	30	44	74	-6
	7	5	19	34	42	76	5	16	31	47	78	-2
	8	13	16	27	43	70	9	16	26	49	75	-4
	9	23	22	24	31	55	18	21	23	38	61	-6
	10	25	20	22	33	55	19	19	22	41	62	-7
Mathematics	3	25	27	32	17	49	13	26	37	25	62	-13
	4	17	39	32	11	43	9	35	40	17	56	-13
	5	18	47	30	5	35	9	40	36	15	51	-16
	6	19	39	32	10	42	11	31	35	23	58	-17
	7	33	30	21	17	37	23	30	24	22	46	-9
	8	37	27	20	16	36	29	27	21	23	44	-8
	9	53	20	16	11	27	38	24	19	19	38	-11
	10	58	20	15	7	22	46	20	18	16	34	-12
Reading	3	47	23	18	12	30	38	22	21	19	40	-10
	4	35	25	23	17	40	28	28	26	18	45	-4
	5	42	24	20	14	34	34	26	23	16	40	-5
	6	39	24	19	18	37	31	23	24	22	45	-8
	7	37	29	24	9	33	30	30	30	10	40	-7
	8	26	25	29	20	49	25	25	31	20	51	-2
	9	38	27	19	16	35	34	23	25	18	43	-9
	10	44	26	22	9	31	40	23	26	11	37	-7

Table 3.2. (continued)

Subject	Grade level	2021 Arkansas					2019 National Norms					Diff. Bench.
		INS	Close	Ready	Exceed.	Met Bench.	INS	Close	Ready	Exceed.	Met Bench.	
Science	3	50	18	14	18	32	42	18	19	22	40	-9
	4	40	23	23	14	37	31	25	26	18	43	-7
	5	39	28	23	10	32	30	26	27	17	44	-12
	6	40	21	25	14	39	30	23	25	21	47	-8
	7	40	21	21	18	38	36	21	22	22	44	-5
	8	42	21	18	18	36	37	21	21	21	43	-6
	9	48	21	19	13	31	42	22	19	16	35	-4
	10	50	22	16	12	28	44	20	20	17	36	-8
ELA	3	49	19	16	17	32	36	21	19	24	44	-12
	4	43	19	18	20	38	36	21	21	23	43	-5
	5	41	25	21	14	35	33	24	22	21	43	-8
	6	39	23	20	18	38	32	22	23	23	46	-8
	7	39	21	20	20	40	33	21	22	23	46	-6
	8	36	20	21	22	43	32	21	22	26	48	-4
	9	43	19	18	21	39	35	17	19	28	47	-9
	10	45	18	18	19	37	37	17	20	26	46	-9
STEM	3					15					20	-5
	4					16					23	-7
	5					9					19	-10
	6					9					18	-9
	7					14					19	-5
	8					13					18	-5
	9					11					17	-6
	10					11					19	-8

Note. INS = In need of support; Exceed. = Exceeding; Bench. = Benchmark. The percentages do not incorporate special scoring rules used by Arkansas for state and federal accountability.

The percentage of students meeting the ACT Readiness Benchmarks is also presented by student group (Table 3.3). Generally, Benchmark attainment was lowest for the English Language Learner group, followed by the Students with Disabilities, African American, and Economically Disadvantaged groups. The percentages may be different than those calculated for state and federal accountability because of the special scoring rules discussed earlier. The methodology used to categorize students by race and ethnicity is discussed in Section 2.1.

For all groups, Benchmark attainment in mathematics decreased across grade levels, particularly for 9th and 10th grade. This is consistent with the national norms, where the percent meeting the mathematics benchmark is 62% for 3rd grade but only 34% for 10th grade – suggesting that the ACT Readiness Benchmarks for mathematics are more difficult to meet as grade level increases.

Table 3.3. 2021 Percentage Meeting ACT Readiness Benchmark, by Group, Subject, and Grade Level

Subject	Grade level	Group							
		Female	Male	African American	Hispanic	White	Econ. Dis.	SWD	ELL
English	3	65	58	42	54	70	54	34	38
	4	66	59	42	56	71	54	33	34
	5	74	65	53	63	76	62	39	41
	6	73	63	50	62	75	61	36	36
	7	82	70	61	73	81	70	46	50
	8	76	65	53	64	77	63	41	39
	9	62	48	33	47	63	45	24	19
	10	63	48	32	46	63	45	24	17
Mathematics	3	49	48	24	41	58	39	23	24
	4	42	44	17	40	52	33	21	25
	5	34	35	13	29	43	26	13	11
	6	42	41	17	37	50	32	18	18
	7	39	36	14	33	46	27	13	7
	8	37	36	12	30	45	26	13	7
	9	27	27	7	21	34	17	9	4
	10	21	22	5	15	27	13	7	3

Table 3.3. (continued)

Subject	Grade level	Group							
		Female	Male	African American	Hispanic	White	Econ. Dis.	SWD	ELL
Reading	3	33	28	13	22	38	22	12	9
	4	43	38	20	32	49	31	16	15
	5	37	31	15	25	42	26	14	7
	6	39	35	17	29	45	28	14	8
	7	37	30	16	28	40	25	12	6
	8	55	43	28	43	57	40	22	17
	9	39	31	15	29	41	26	13	7
	10	35	27	13	25	37	22	12	5
Science	3	31	32	11	22	41	23	14	8
	4	35	38	14	28	46	27	15	11
	5	31	34	11	23	41	23	16	8
	6	38	39	16	31	48	29	16	8
	7	39	38	16	33	47	29	16	8
	8	37	35	13	29	45	27	15	6
	9	32	30	11	23	39	22	12	5
	10	29	27	9	20	36	19	11	4
ELA	3	36	28	14	26	39	23	13	11
	4	42	34	16	31	46	28	13	11
	5	39	30	15	29	42	26	12	7
	6	44	33	17	32	46	29	12	8
	7	48	32	19	35	47	30	13	7
	8	51	36	21	36	52	33	16	9
	9	46	32	18	32	46	29	13	7
	10	44	30	15	30	45	27	13	6
STEM	3	14	17	3	8	21	9	6	1
	4	14	17	3	10	21	9	6	2
	5	8	10	1	5	12	5	3	<1
	6	8	10	1	5	12	5	3	1
	7	13	14	3	8	18	7	4	<1
	8	12	14	2	8	17	7	4	<1
	9	11	12	2	6	15	6	4	<1
	10	10	12	2	6	15	5	4	<1

Note. Econ. Dis. = Economically Disadvantaged; SWD = Students with Disabilities; ELL = English Language Learner. The percentages do not incorporate special scoring rules used by Arkansas for state and federal accountability.

3.2 Achievement Trends From 2016 to 2021

ACT Aspire has been used as the Arkansas state assessment since spring 2016. In this section, we address the question, “How has the achievement of Arkansas students changed over time”?

Appendix [Table A.2](#) presents summary measures of achievement for each year from 2016 to 2021, for each subject and grade level. The summary measures include the mean score and the percentage meeting the ACT Readiness Benchmark.

Generally, improving score trends were observed from 2016 to 2019, followed by declines from 2019 to 2021. The declines from 2019 to 2021 may be attributed to the COVID-19 pandemic. In later sections of this report, we present additional analyses of Arkansas’ scores from 2019 and 2021 to better understand possible impacts of the pandemic.

Across the 48 subject and grade level combinations (not including composite), mean scores increased from 2016 to 2019 in 41 cases and decreased in 7 cases. The largest improvements in mean scores were observed for 8th grade mathematics (+1.7 score points), 9th grade science (+1.6 score points), and 8th grade ELA (+1.5 score points). The largest decreases in mean scores were observed for 6th grade mathematics (-0.9 score points), science (-0.9 score points), and STEM (-0.9 score points).

Similarly, the percentage meeting the ACT Readiness Benchmark increased from 2016 to 2019 in 40 cases and decreased in 8 cases. The largest improvements in Benchmark attainment were observed for 4th grade ELA (+12%), 8th grade mathematics (+10%), and 8th grade ELA (+10%). The largest decreases in Benchmark attainment were observed for 6th grade science (-7%), 6th grade STEM (-3%), and 6th grade reading (-2.5%).

From 2019 to 2021, there were decreases in mean scores and percentage meeting the ACT Readiness Benchmark for all grade levels and subjects. The largest declines in mean scores were observed for 8th–10th grade mathematics (-2.3, -2.1, and -2.2 score points, respectively). The largest declines in percentage meeting the ACT Readiness Benchmarks were observed for mathematics for 5th grade (-15%), 3rd grade (-13%), and 6th grade (-12%).

For 2021, the overall (across grades 3–10) percentage meeting the Benchmark (scoring at the Ready or Exceeding levels) was 65% for English, 36% for reading, 38% for ELA, 36% for mathematics, 34% for science, and 12% for STEM. Trends in Benchmark and Readiness Level attainment from 2016 to 2021 are summarized in Figures 3.1–3.5. The figures show the percentage of students across grades 3–10 performing at each ACT

Readiness Level. For all subjects, there is improvement in Benchmark attainment from 2016 to 2019. Most of the improvement occurred from 2016 to 2017, but notable improvement also occurred from 2018 to 2019 for mathematics and ELA. For all subjects, there is a decline in Benchmark attainment from 2019 to 2021, with the largest decline observed for mathematics.

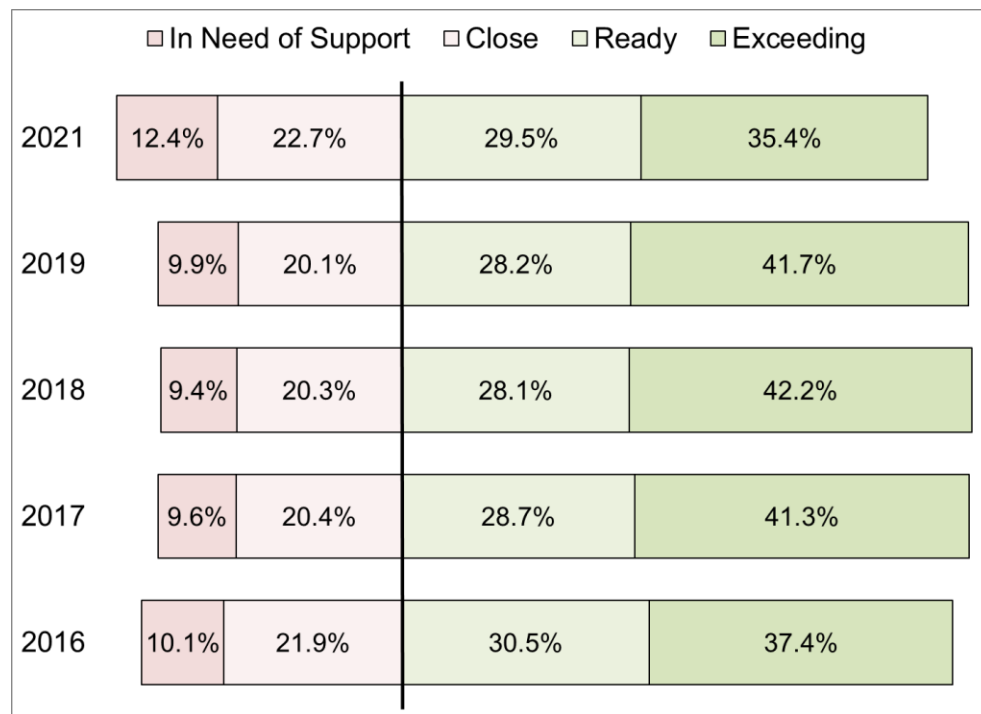


Figure 3.1. Overall percentage meeting ACT Readiness Levels for English, by year

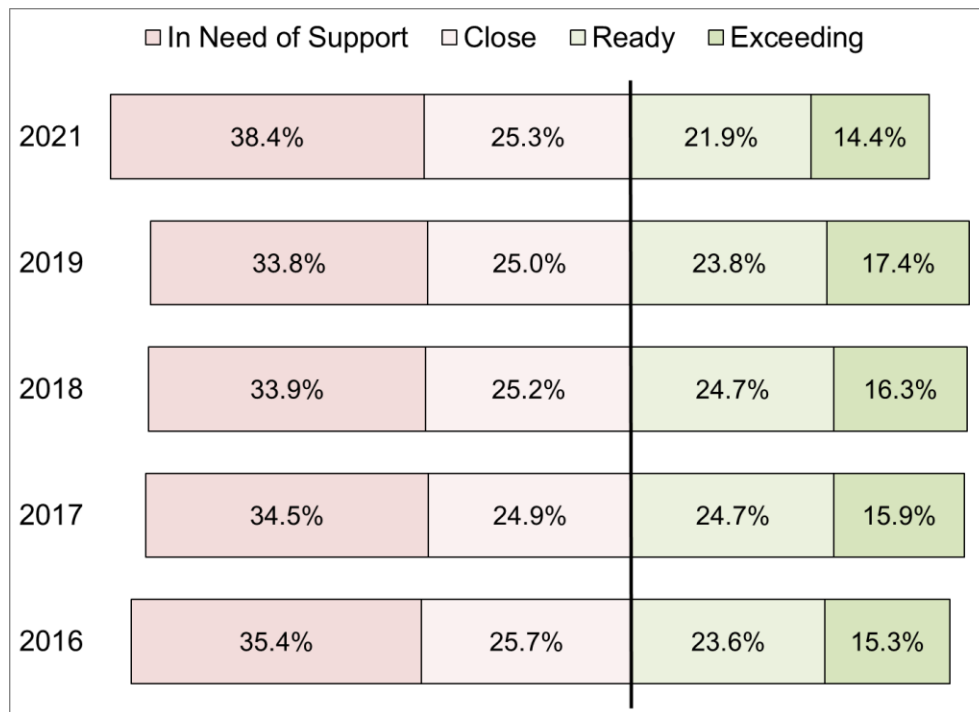


Figure 3.2. Overall percentage meeting ACT Readiness Levels for reading, by year

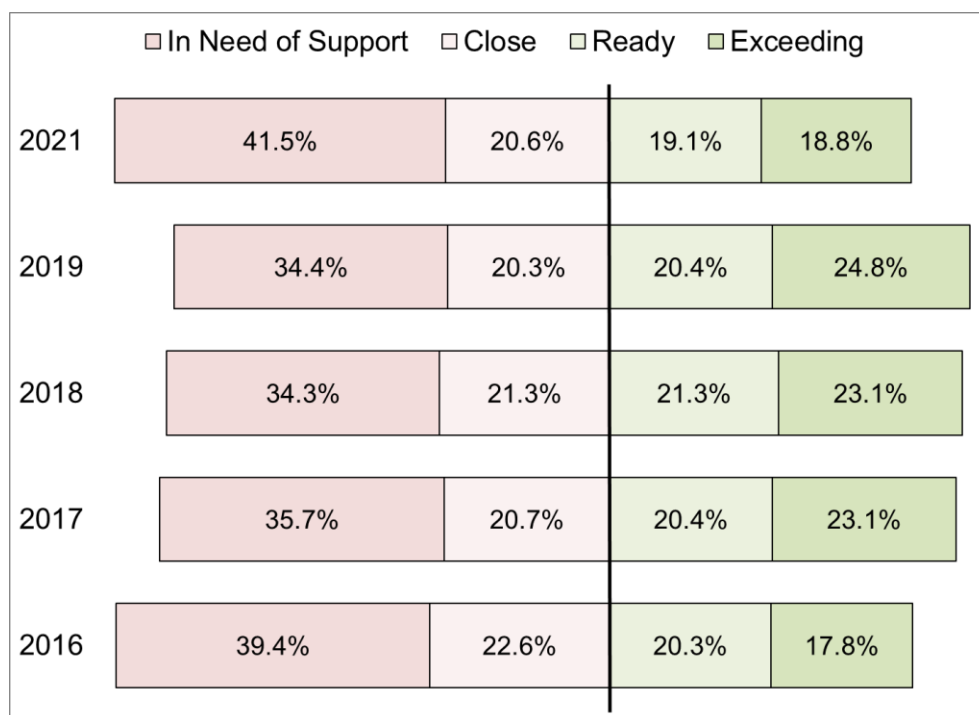


Figure 3.3. Overall percentage meeting ACT Readiness Levels for ELA, by year

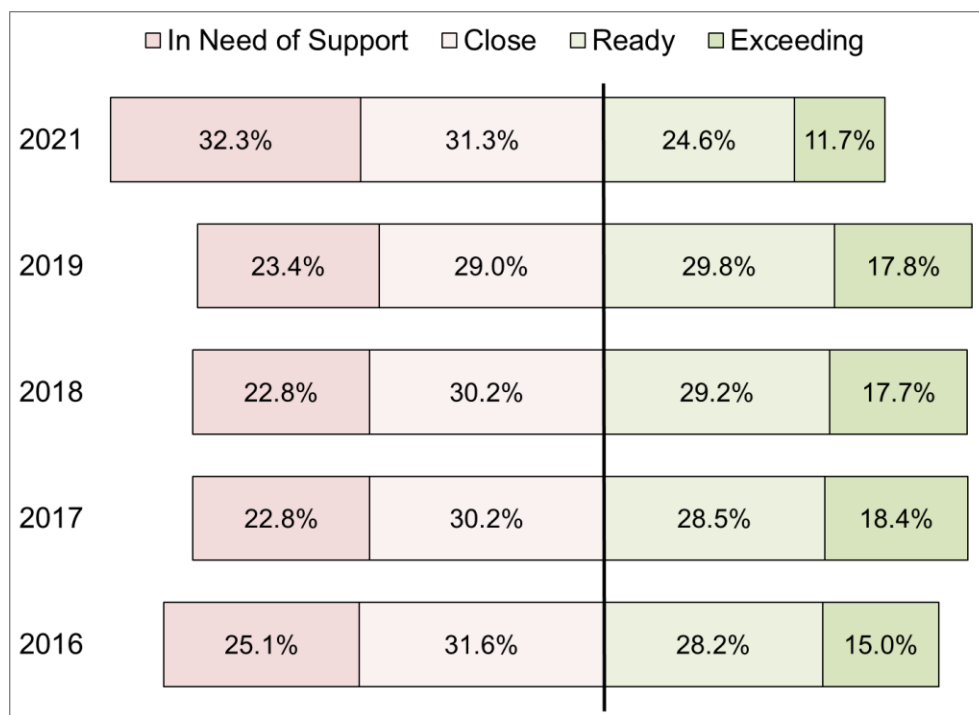


Figure 3.4. Overall percentage meeting ACT Readiness Levels for mathematics, by year

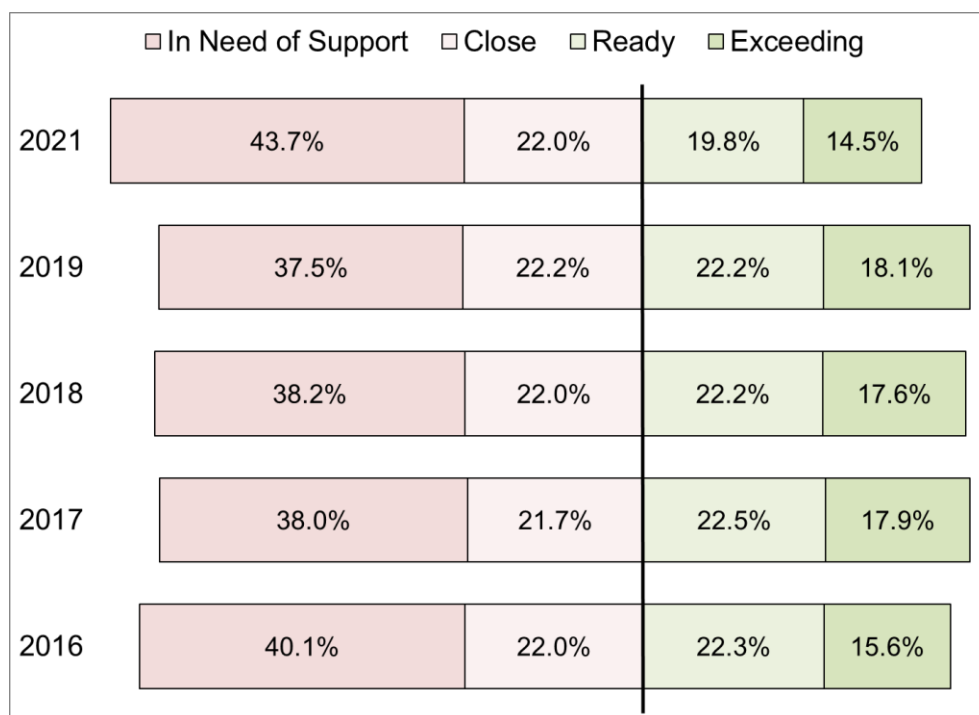


Figure 3.5. Overall percentage meeting ACT Readiness Levels for science, by year

3.3 Understanding the Impact of the COVID-19 Pandemic: Comparing Performance from 2019 and 2021

The COVID-19 pandemic caused unprecedented disruptions to the educational system in Arkansas and across the United States. At the onset of the pandemic in March 2020, schools were forced to replace onsite instruction with virtual instruction. During the 2020–2021 academic year, there were three student instructional options: onsite/traditional learning, hybrid/blended learning, or virtual/remote learning.

In Section 3.2, we reported that average scores and percentage meeting the ACT Readiness Benchmarks decreased substantially from 2019 to 2021. From that analysis, it is not clear how much of the change in performance was due to changes in test participation and the tested population, and how much was due to pandemic-related performance declines. In this section, we present a formal analysis of performance in spring 2019 (before the pandemic) and spring 2021 (during the pandemic) and account for changes in test participation and the tested population across years. Because the same primary test forms were used in 2019 and 2021, we can be more confident that changes across the two cohorts are not confounded by test form effects. The analysis focuses on the change in average (mean) test scores.

3.3.1 Data and Methods

ACT Aspire and 11th grade ACT test scores from 2019 and 2021 were matched to student enrollment and demographic data provided by the Arkansas Department of Education. For grades 5–11, the data were also matched to prior ACT Aspire Summative test scores (from spring 2017 for the 2019 cohort, from spring 2019 for the 2021 cohort). Finally, the data were matched to instructional option data provided by the Arkansas Department of Education. As described later, the instructional option data are used to group students according to their mode of learning (onsite, hybrid, or virtual) during the 2020–2021 academic year.

Analysis was conducted for each subject area (including ELA, STEM, and Composite) and grade level (3–11). Figure 3.5 shows the sample sizes for the analysis of Composite test scores for the 2021 cohort. “Population” represents the number of students in the Arkansas population with demographic data. The “Status” sample represents the number of students in the Arkansas population and with a Composite test score. The “Growth” sample represents the subset of students in the Status sample who also had a prior ACT Aspire Composite test score. As described more in a later section of this report, the “Mode of Learning” sample is the subset of students who also had mode of learning data (Onsite, Hybrid, or Virtual) reported for 2020–2021.

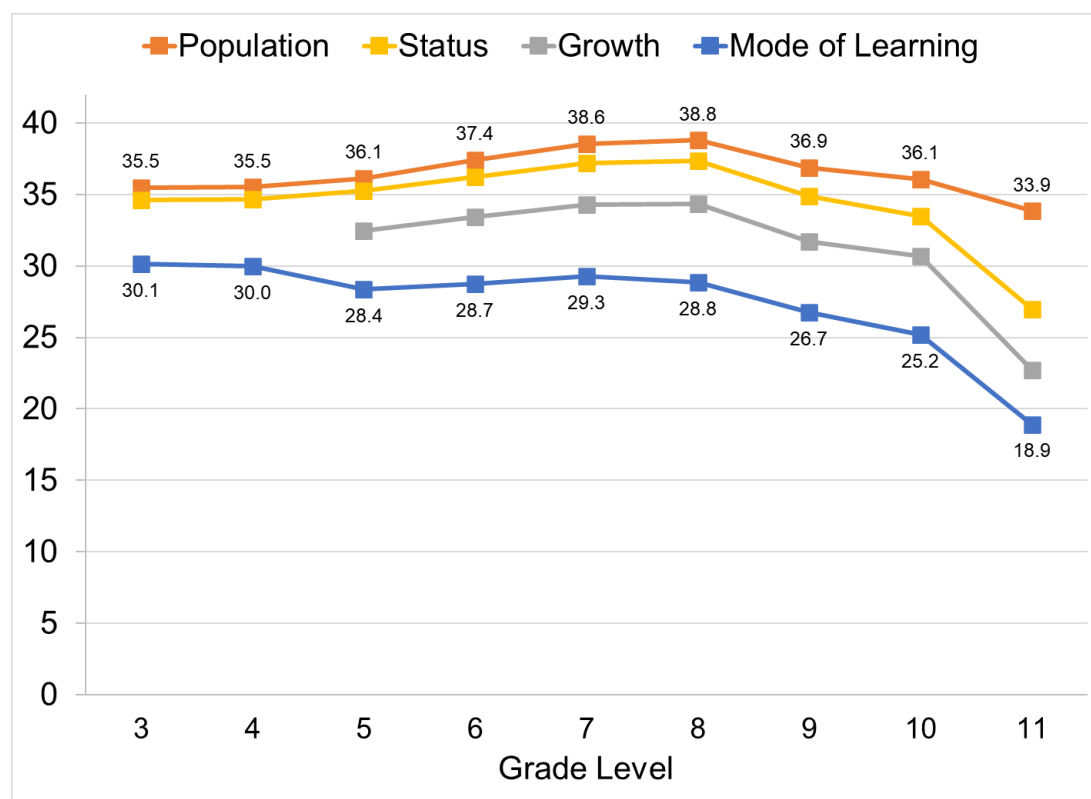


Figure 3.5. Sample sizes for analysis of COVID-19 impacts on test scores

To ensure that the two cohort groups are similar on demographic characteristics, propensity score weighting (Austin, 2011) was used to weight each group to be similar to the pooled data set, which combines the 2019 and 2021 data. The procedure uses logistic regression to estimate each student’s propensity for being in each cohort based on their gender, race/ethnicity, disability status, economic disadvantage status, and English learner status. For grades 5–11, prior ACT Aspire summative test scores (from spring 2017 for the 2019 cohort and from spring 2019 for the 2021 cohort) were used to ensure that the two cohort groups (2019 and 2021) did not differ on prior academic achievement.¹

¹ ACT Aspire scale scores are comparable across test forms through rigorous equating processes. Further, because the same primary test forms were administered in both 2019 and 2021, the changes in average test scores are not likely attributed to test form differences. For grades 5–11, prior ACT Aspire test scores from 2017 and 2019 were also used in the analysis, with a post-hoc adjustment applied to these prior test scores to reduce possible random errors. Because the same ACT Aspire Interim test forms were used across years, Interim test scores can be used as anchors to adjust average Summative test scores across years. The post-hoc adjustment quantity was derived using linear regression with ACT Aspire Interim test scores as predictors of ACT Aspire Summative test scores from 2017 and 2019. For example, if Summative test scores were 0.2 points lower in 2017 relative to 2019 (after adjusting for interim test scores), the scores from 2017 were adjusted +0.2 score points.

Using the weighted data, hierarchical linear regression models were fit to estimate the difference in average test scores from 2019 to 2021. The models included a random intercept for the school effect and included the same covariates that were used for the propensity score weighting model. The general form of the regression model was:

$$\text{Test Score} = \text{Cohort Year} + \text{Covariates} + \text{School Effect}$$

The regression model produced an estimate of the difference in average test score from 2019 to 2021, adjusted for changes in the tested population. We refer to this as the adjusted score difference. The adjusted score difference can be expressed in three ways:

- 1) On the original test score scale.
- 2) In standard deviation units (e.g., d_{sd} = adjusted score difference divided by the standard deviation of the test score).
- 3) In average yearly gain units (e.g., d_{gain} = adjusted score difference divided by the estimated score gain normally observed over one academic year). To estimate the score gain normally observed over one academic year, we considered pre-pandemic data from Arkansas (one year gains observed in 2017, 2018, and 2019 for grades 4–10). A linear regression model was used to estimate the mean gain for each grade level and subject, smoothing out differences across grade levels and extrapolating for 3rd grade. For the ACT test (11th grade), the average pre-pandemic gain is based on students who took the ACT as part of school-day testing in 10th grade and took the ACT gain in 11th grade 10–14 months later.

The hierarchical linear regression models were also used to estimate the adjusted score difference for different student groups, including:

- Male and female students
- Racial/ethnic groups: African American, Asian, Hispanic, Native American/Native Hawaiian, White, and Two or more races
- Students with disabilities and students without disabilities
- Economically disadvantaged students and non-economically disadvantaged students
- English language learners (ELL) and non-ELL students

Group-specific estimates were obtained by fitting the hierarchical linear regression model with interactions between the cohort (year) and the group indicator.

3.3.2 Total Group Results

The adjusted Composite score differences provide an overall estimate of the pandemic's impact on ACT Aspire and ACT test scores. Figure 3.6 provides the estimated Composite score declines from 2019 to 2021 by grade level. The estimate based on the growth sample that accounts for prior test scores, student demographics, and school effects is reported by the gray line. The estimate based on the Status sample that accounts for student demographics and school effects is reported by the blue line. The growth sample estimate is only available for grades 5–11 whereas the Status sample estimate is available for grades 3–11. For grades 6–11, the decline in Composite score is more severe for the growth sample, suggesting that not accounting for prior test scores may understate the pandemic's impact on ACT Aspire and ACT test score. For 11th grade, note that the score declines are expressed on the ACT Composite score scale, which is different than the ACT Aspire score scales. Figure 3.6 shows that Composite score declines were observed for all grade levels.

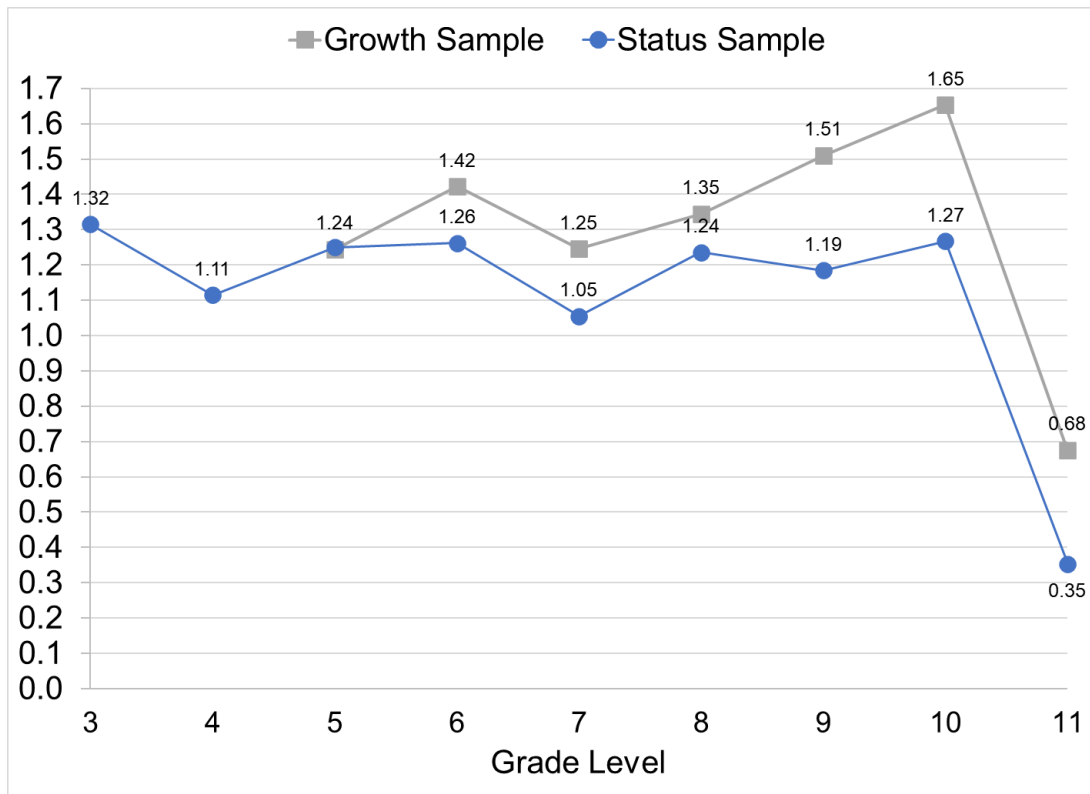


Figure 3.6. Decline in average Composite test scores from 2019 to 2021

Figure 3.7 provides the estimated Composite score declines, in standard deviation units, by grade level. The estimated score declines in Figure 3.7 can be interpreted as standardized effect sizes. Score declines reported in this fashion allow us to interpret the score declines relative to the distribution of test scores at each grade level. For example, $d_{sd}=0.10$ is comparable to a percentile rank decrease of four percentile rank units (e.g., moving from the 50th to the 46th percentile). Generally, the standardized score declines were larger for lower grade levels. For 3rd grade, Composite scores declined by 0.25 standard deviation units, which is like moving from the 50th to the 40th percentile of scores.

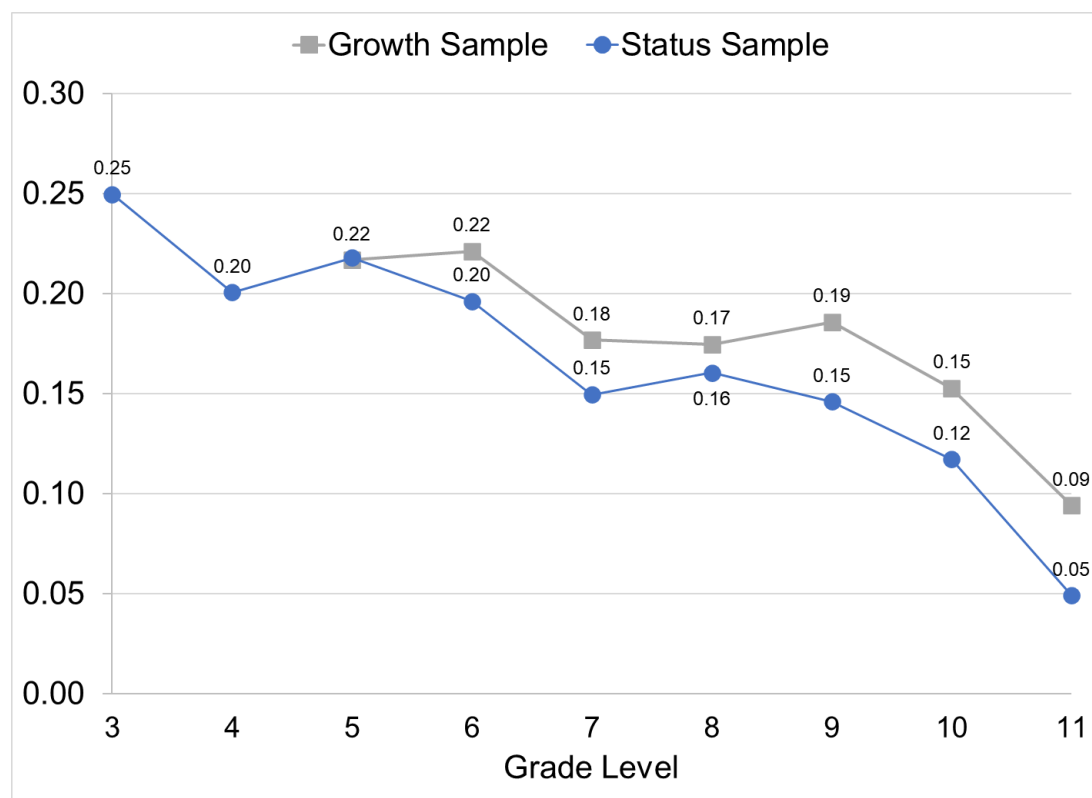


Figure 3.7. Decline in average Composite test scores, in standard deviation units, from 2019 to 2021

Figure 3.8 provides the estimated Composite score declines, in average annual gain units, by grade level. The estimated score declines in Figure 3.8 allow us to interpret the score declines relative to the yearly score gain that is normally observed for each grade level. For example, $d_{gain}=0.50$ indicates that the score decline is comparable to half of the score gain that is normally observed in one year. The most extreme result was observed for 9th grade Composite scores ($d_{gain}=1.07$), suggesting that the decline in scores from 2019 to 2021 is comparable to approximately one year of typical score gain. It is worth noting that d_{gain} is based on estimates for both the numerator (the

adjusted score difference) and the denominator (the average score gain from 8th to 9th grade), so is especially prone to being over- or under-estimated. Reflecting this uncertainty, a 90% confidence interval for d_{gain} in this case is [0.78, 1.70].

Relative to the average yearly gains, the score declines were smaller for lower grade levels, with the exception of the 10th grade ACT Aspire and 11th grade ACT test results. Contrasting Figures 3.7 and 3.8 we see that the relationship of grade level and score decline depends on which score decline metric is used. Relative to the distribution of test scores at each grade level, the score declines are less severe for later grade levels. But relative to annual score gains, the score declines are generally more severe for later grade levels. This apparent paradox is caused by the variability (standard deviation) of test scores increasing with grade level, while average score gains tend to decrease with grade level.

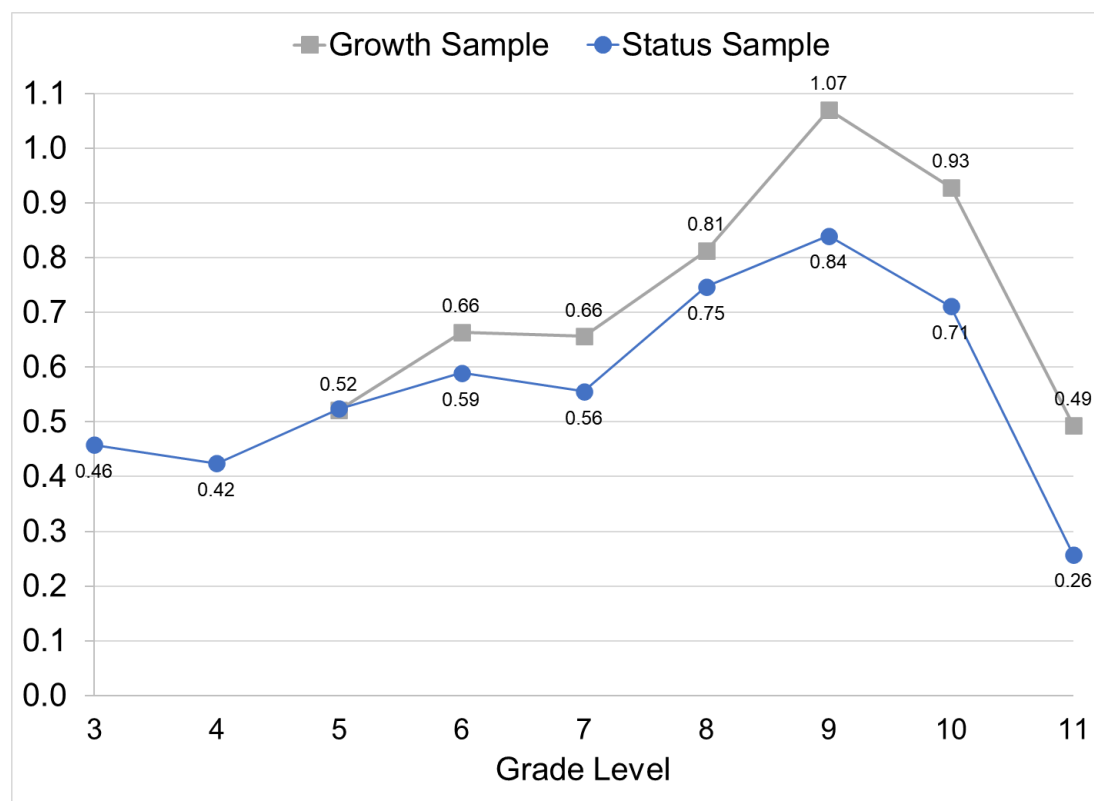


Figure 3.8. Decline in average Composite test scores, in average annual gain units, from 2019 to 2021

Figure 3.9 provides the estimated score declines, in standard deviation units, by subject and grade level. Note that the estimates are based on the Status sample for 3rd and 4th grade and the Growth sample for grades 5–11. Therefore, the results for 3rd and 4th grade may understate the pandemic’s impact. Because they are based on different

statistical models, the lines between 4th and 5th grade are not connected. For grades 3–10, we see that the relative score declines were much larger in Math relative to the other subject areas. The score declines were generally more severe for English and science than for Reading. For all grade levels and all subjects, average scores declined from 2019 to 2021.

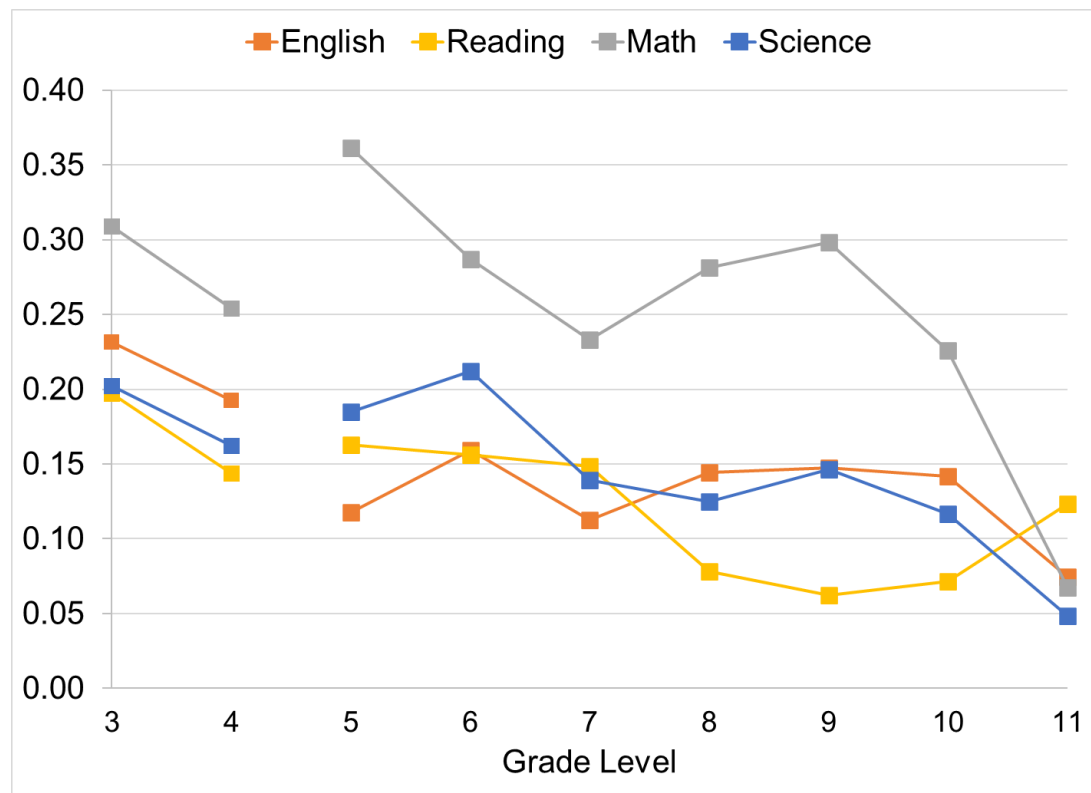


Figure 3.9. Decline in average subject test scores, in standard deviation units, from 2019 to 2021

3.3.3 Results by Student Group

The decline in ACT Aspire and ACT test scores from 2019 to 2021 was also examined for each student group. Figure 3.10 shows the decline in average Composite scores, by grade level, for African American, Hispanic, and White racial/ethnic groups. For grades 5–11, the score declines were less severe for the African American and Hispanic groups relative to the White group. For 3rd and 4th grade, the estimated score declines do not account for prior test scores and so are based on the Status sample rather than the Growth sample. For 3rd and 4th grade, the estimated score declines are more severe for the African American group relative to the Hispanic and White groups.

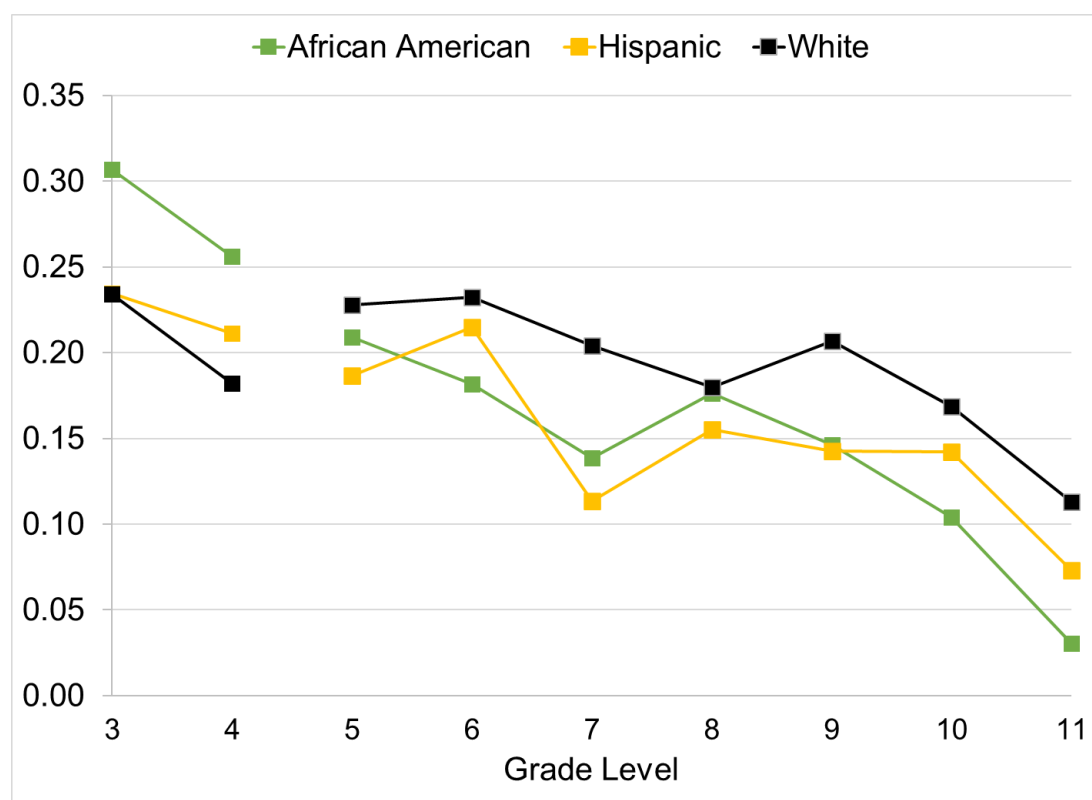


Figure 3.10. Decline in average Composite test scores by race/ethnicity, in standard deviation units, from 2019 to 2021

Figure 3.11 shows the decline in average Composite scores, by grade level and disability status. The score declines were consistently less severe for the student with disability group relative to students without disabilities. Similarly, Figure 3.12 shows that the score declines were consistently less severe for the English learner group relative to the not English learner group.

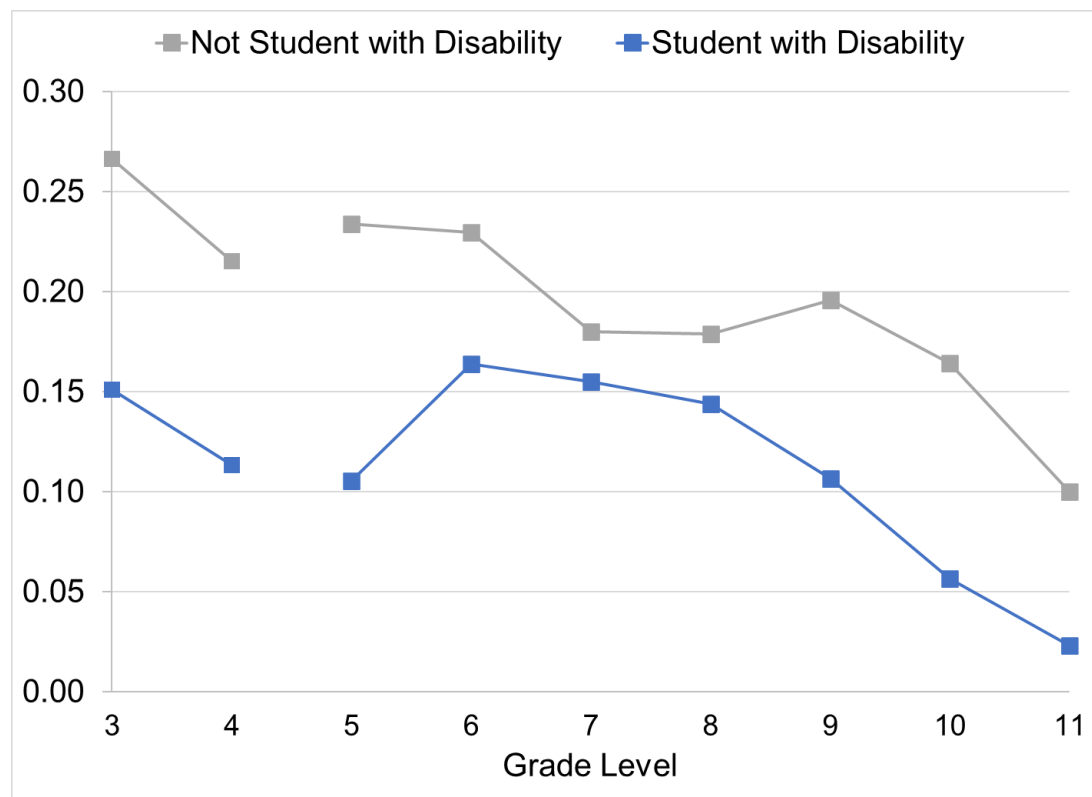


Figure 3.11. Decline in average Composite test scores by disability status, in standard deviation units, from 2019 to 2021

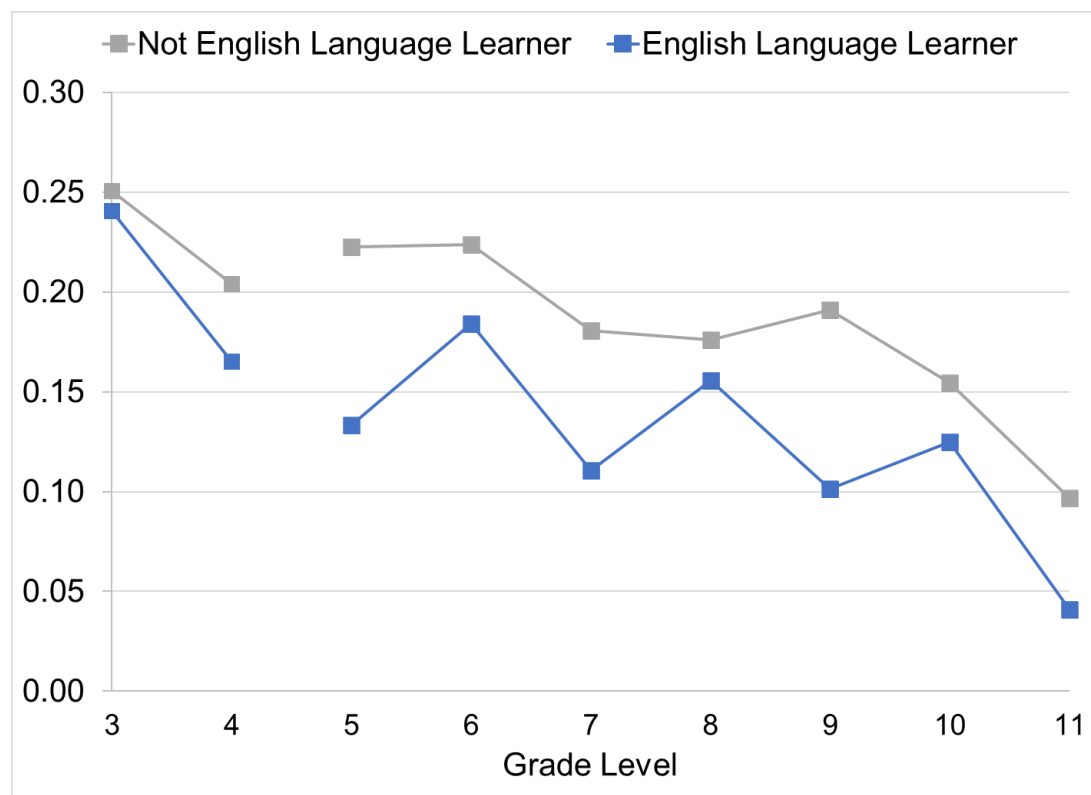


Figure 3.12. Decline in average Composite test scores by English learner status, in standard deviation units, from 2019 to 2021

Score declines were also examined by gender and by economic status (results not presented). Score declines were generally more severe for female students relative to male students, but the differences were very small. The differences in score declines by economic were small and not consistent across grade levels.

3.3.4 Summary of COVID-19 Impact on Performance

- After controlling for changes in the tested population from 2019 to 2021, average ACT Aspire scores declined for all subjects and grade levels, suggesting that the COVID-19 pandemic has had a negative effect on student learning.
- Score declines, expressed relative to the standard deviation of test scores, suggest that the severity of the pandemic's impact increases with grade level; score declines, expressed relative to average annual score gains, suggest that the severity of the pandemic's impact decreases with grade level.
- The pandemic-related score declines are most severe for math and least severe for reading.

- For grades 5–11, the score declines were less severe for African American and Hispanic students relative to White students.
- The score declines for students with disabilities or English language learners were less severe than the score declines observed for those without disabilities or students who are not English language learners.

3.4 Differences in Performance by 2020–2021 Instructional Option: Comparing Onsite, Hybrid, and Virtual Modes of Learning

In Section 3.3 we reported that average scores declined from 2019 to 2021 for all subjects and grade levels. In this section, we present additional analyses examining whether performance in spring 2021 varied by mode of learning. Analysis was conducted for each subject area and grade level (3–11).

3.4.1 Data and Methods

The data used for this analysis are a subset of the Growth sample (for grades 5–11) and the Status sample (for 3rd and 4th grade). Figure 3.5 shows the Composite score sample sizes by grade level. For students tested in 2021, the Composite sample size for the mode of learning analysis ranged from 30.1 thousand for 3rd grade to 18.9 thousand for 11th grade.

During the 2020–2021 academic year, data were collected on students' instructional option: Onsite, Hybrid, or Virtual. This set of data was collected at four time points: November 15, 2020 (Cycle 3); February 15, 2021 (Cycle 5); April 15, 2021 (Cycle 6); and June 15, 2021 (Cycle 7). Approximately 95% of students had mode of learning data reported for all four cycles.

Students were categorized as follows:

- **All Onsite** - Students who were classified as Onsite at all four cycles.
- **All Hybrid** - Students who were classified as Hybrid at all four cycles.
- **All Virtual** - Students who were classified as Virtual at all four cycles.
- **Hybrid then Onsite** - Students who were classified as Hybrid for Cycle 3 or Cycle 3 and 5 and then classified as Onsite for the later cycles.
- **Virtual then Onsite** - Students who were classified as Virtual for Cycle 3 or Cycle 3 and 5 and then classified as Onsite for the later cycles.

- **Other: 75% or more in-person** - Students who did meet criteria for the other categories and who had a weighted average of 75% or more in-person across cycles. (Onsite is considered 100% in-person, Hybrid is 50% in-person, and Virtual is 0% in-person).
- **Other: 38–63% or more in-person** - Students who did meet criteria for the other categories and who had a weighted average of 38–63% in-person across cycles.
- **Other: 25% or less in-person** - Students who did meet criteria for the other categories and who had a weighted average of 25% or less in-person across cycles.
- **Missing** - Students who were missing instructional option data for one or more cycles.

Note that the accuracy of the classifications is limited because of the nature and timing of the mode of learning data collected. For example, a student who began the 2020–2021 academic year learning in the Virtual mode but switched to Onsite before the first data collection date (November 15) could be classified as “All Onsite,” even though they spent some time learning Virtually.

Table 3.4 provides the relative frequency of each category by grade level. The percentage of students learning purely onsite decreased with grade level, from a high of 65% for 3rd grade to a low of 48% for 11th grade. Conversely, the percentage of students learning purely in hybrid mode increased from 3rd grade (9%) to 11th grade (12%) and the percentage of students learning purely in virtual mode increased from 3rd grade (12%) to 11th grade (21%).

Table 3.4. Percent in each mode of learning category, by grade level

Category	Grade Level								
	3	4	5	6	7	8	9	10	11
All Onsite	65	64	64	61	60	58	56	50	48
All Hybrid	9	9	9	9	10	10	10	12	12
All Virtual	12	13	13	14	14	15	16	18	21
Hybrid then Onsite	1	1	1	1	1	2	1	3	3
Virtual then Onsite	4	4	4	5	4	5	4	4	4
Other: $\geq 75\%$ in-person	1	1	1	2	2	2	3	3	4
Other: 38–63% in-person	1	1	1	2	2	2	3	3	3
Other: $\leq 25\%$ in-person	0	1	1	1	1	1	1	1	1
Missing	6	6	5	5	5	5	5	5	4

Table 3.5 summarizes participation in each mode of learning category by student group. African American (27%) and Asian (34%) students were considerably more likely to participate in Virtual instruction relative to other racial/ethnic groups, while Hispanic (65%), Native American/Hawaiian (64%), and White (61%) students were more likely to participate in Onsite instruction. English language learners (67%) were also more likely to participate in Onsite instruction.

Table 3.5. Percentage in each mode of learning category, by student group

Group	Model of Learning Category				
	All Onsite	All Hybrid	All Virtual	Other	Missing
Total	59	10	15	11	5
African American	46	7	27	15	5
Asian	44	8	34	11	3
Hispanic	65	8	12	10	5
Native American/Hawaiian	64	10	9	6	11
Two or More Races	59	8	14	12	7
White	61	12	12	10	5
Student with Disability	60	10	15	9	6
Economically Disadvantaged	60	9	16	12	3
English Language Learner	67	8	9	9	7

An analysis of mode of learning differences is simplified when there are fewer groups and the groups are well-defined. Because the majority of students were classified as All Onsite, All Hybrid, or All Virtual (ranging from 87% for 3rd grade to 81% for 10th and 11th grades), and because the percentage of students not classified into these groups was similar across student groups, the analysis was limited to a comparison of those three groups.

To compare test scores for each mode of learning group relative to a pre-pandemic baseline group, we included the total 2019 sample as the reference group. Propensity score weighting was used to weight each of the four groups to be similar to the pooled data set, which combines the 2019 data and the three model of learning groups from 2021. Logistic regression was used to estimate each student's propensity for being in each group based on their gender, race/ethnicity, disability status, economic disadvantage status, and English learner status. For grades 5–11, prior ACT Aspire Summative test scores were used to ensure that the four groups did not differ on prior academic achievement.

Using the weighted data, hierarchical linear regression models were fit to estimate the difference in average test scores across the four groups. The models included a random intercept for the school effect and included the same covariates that were used for the propensity score weighting model. The general form of the regression model was:

$$\text{Test Score} = \text{Group} + \text{Covariates} + \text{School Effect}$$

The regression model produced an estimate of the difference in average test score for each mode of learning group, relative to the 2019 reference group, and adjusted for changes in the tested population. As with the earlier analysis of pandemic-related score declines, adjusted score differences can be expressed on the original test score scale, in standard deviation units (e.g., d_{sd} = adjusted score difference divided by the standard deviation of the test score), or in average yearly gain units (e.g., d_{gain} = adjusted score difference divided by the estimated score gain normally observed over one academic year).

The hierarchical linear regression models were also used to estimate the adjusted score difference for different student groups, including:

- Male and female students
- Racial/ethnic groups: African American, Asian, Hispanic, Native American/Native Hawaiian, White, and Two or more races
- Students with disabilities and students without disabilities
- Economically disadvantaged students and non-economically disadvantaged students
- ELL students and non-ELL students

Group-specific estimates were obtained by fitting the hierarchical linear regression model with interactions between group and the group indicator.

3.4.2 Total Group Results

The adjusted Composite score differences between each mode of learning group and the 2019 reference group are provided in Figure 3.13. For grades 5–11, the estimates account for prior test scores, student demographics, and school effects. For 3rd and 4th grade, the estimates account for student demographics and school effects. Because different models are used for grades 3–4 and 5–11, the lines connecting 4th and 5th grade are not presented. For 11th grade, note that the score declines are expressed on the ACT Composite score scale, which is different than the ACT Aspire score scales.

For all grade levels, average scores for the three 2021 mode of learning groups were below the scores for the 2019 pre-pandemic group. And, the score declines were most severe for the Virtual group, followed by the Hybrid group, and finally the Onsite group. The differences between Virtual and Hybrid are generally larger than the differences between Hybrid and Onsite. The differences between the Onsite group and the reference group are large, showing that even students in the Onsite group suffered score declines relative to the pre-pandemic group. Generally, the performance gap for students who learned Virtually is smaller for higher grade levels.

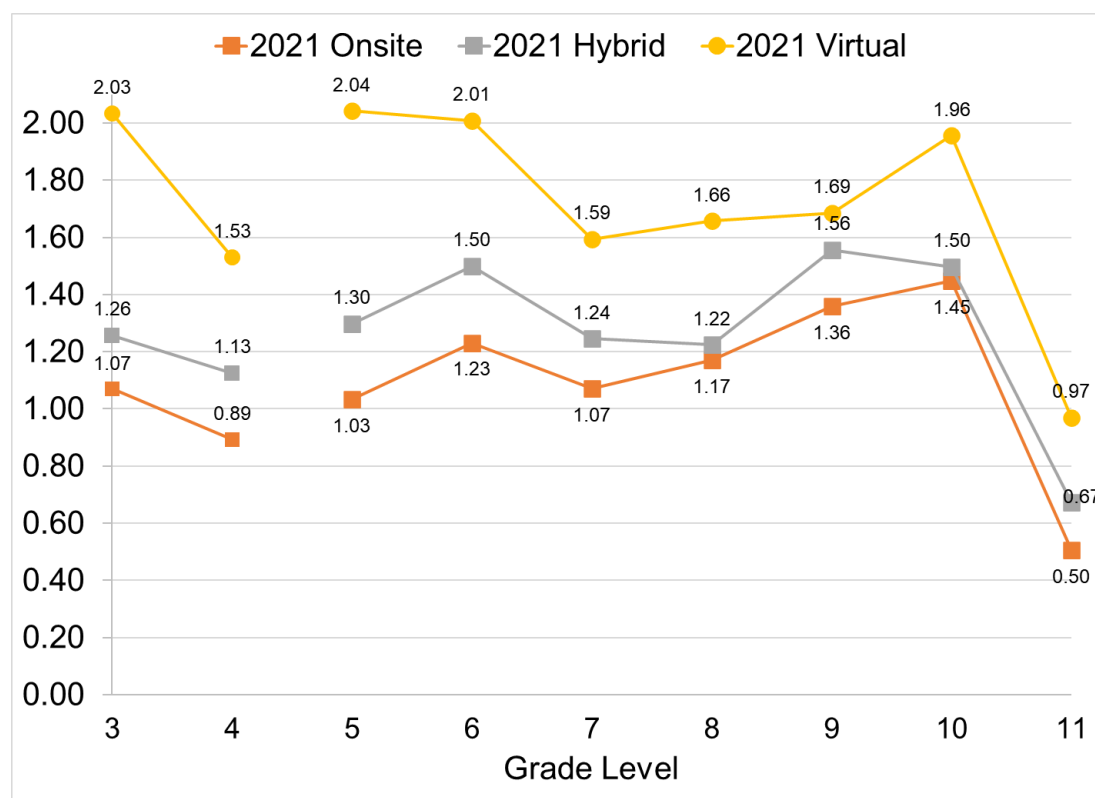


Figure 3.13. Decline in average Composite test scores relative to 2019, by mode of learning group

The adjusted Composite score differences between each mode of learning group and the 2019 pre-pandemic group, expressed in standard deviation units, are provided in Figure 3.14. Relative to the distribution of test scores at each grade level, we see that the differences between the 2021 mode of learning groups and the 2019 group tend to become smaller for higher grade levels.

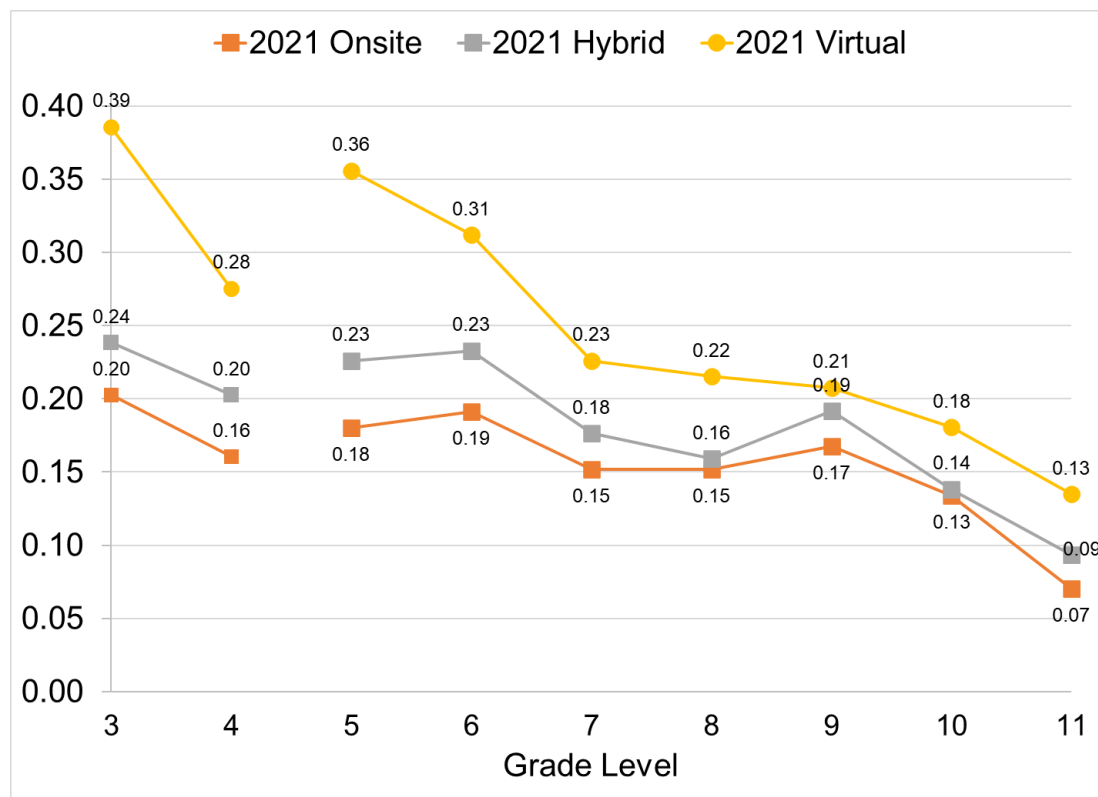


Figure 3.14. Decline in average Composite test scores (in standard deviation units) relative to 2019, by mode of learning group

The adjusted Composite score differences between each mode of learning group and the 2019 pre-pandemic group, expressed in annual gain score units, are provided in Figure 3.15. Relative to the average annual gain in scores, the differences between the 2021 mode of learning groups and the 2019 group tend to become larger for higher grade levels. Note that the 11th grade results do not follow this trend, but are based on the ACT Composite score, which is not on the same scale as the ACT Aspire Composite score.

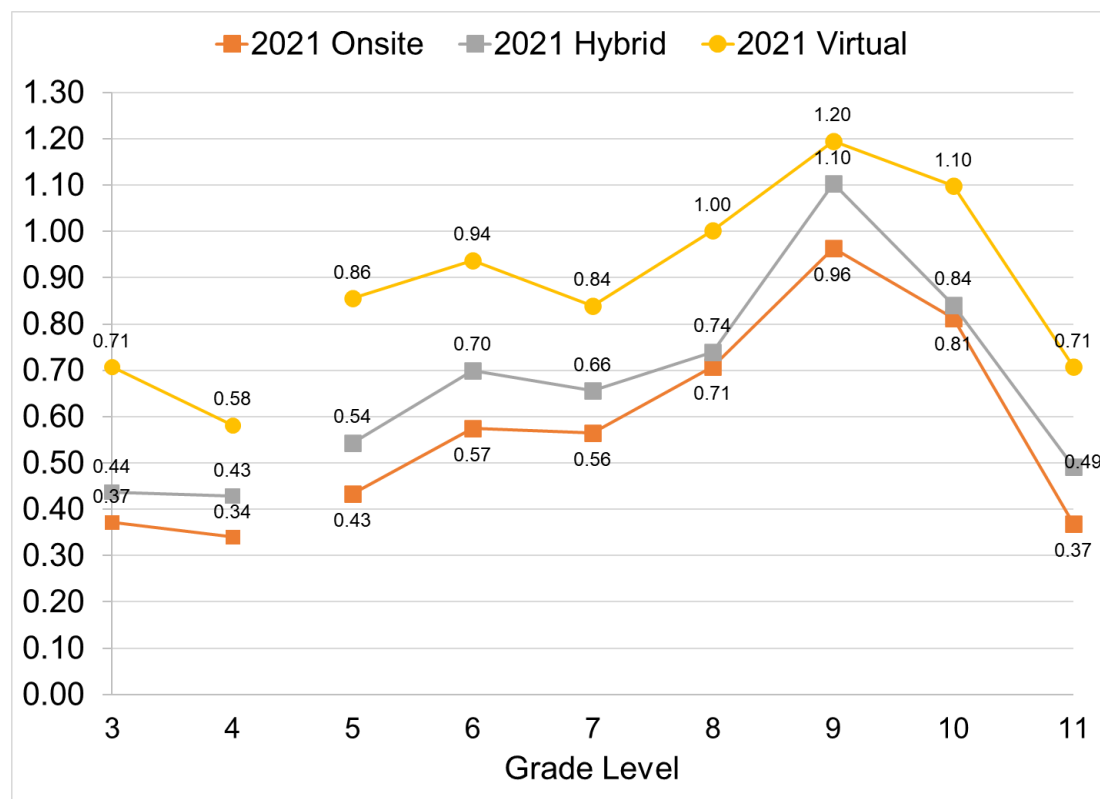


Figure 3.15. Decline in average Composite test scores (in annual gain score units) relative to 2019, by mode of learning group

The grade level trends shown in Figures 3.13, 3.14, and 3.15 are similar to the trends for the overall analysis of score declines (Figures 3.6, 3.7, and 3.8). The estimates for the three mode of learning groups represent a disaggregation of the estimates for the overall 2021 group. For example, 5th grade ACT Aspire Composite scores decreased by 1.24 points in 2021 relative to 2019 (Figure 3.6). The score decrease varied by mode of learning group (Figure 3.13) with average score declines of 1.03 for Onsite, 1.30 for Hybrid, and 2.04 for Virtual.

Another way to understand differences across groups is to compare average gain scores. Gain scores are only available for grades 5–10 because prior ACT Aspire scores are required to calculate the gain score. (Gain scores are not available for 11th grade because ACT test scores are not reported on the same vertical scale as ACT Aspire test scores.) Figure 3.16 shows the average gain in Composite test scores for each mode of learning group relative to the 2019 pre-pandemic group (blue line). For each grade level and mode of learning group, the gain scores are substantially lower than what was observed for the 2019 group. The differences are most pronounced for students in the Virtual group.

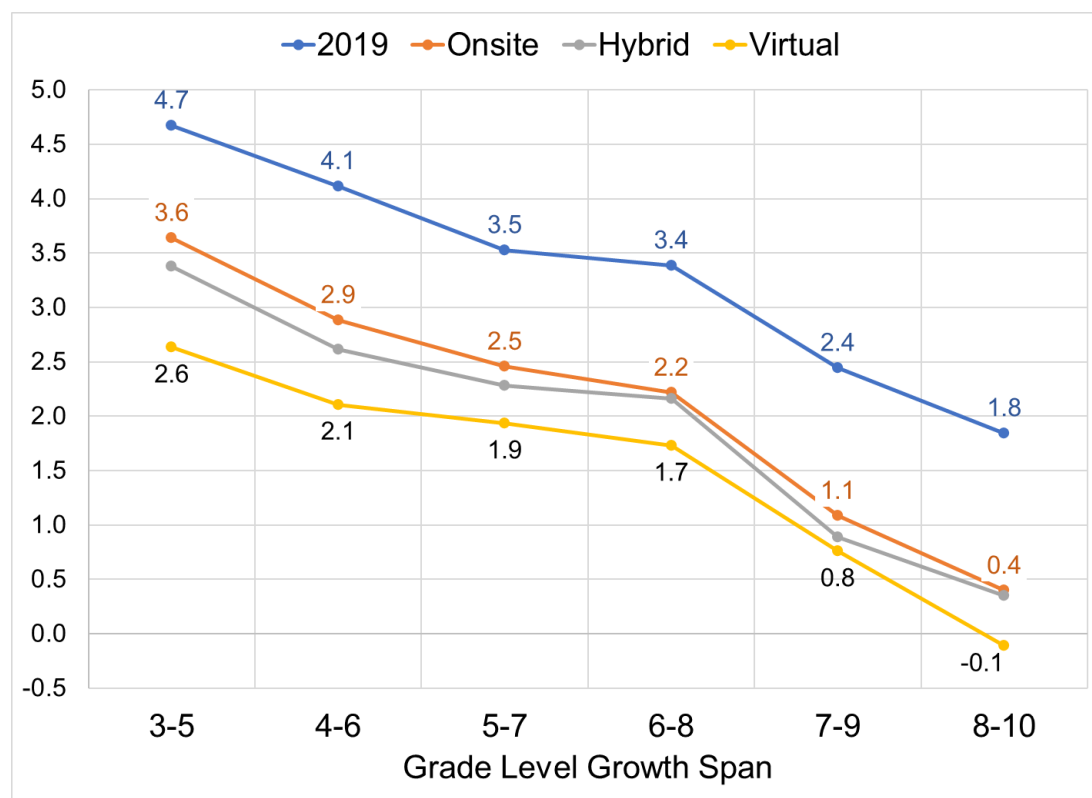


Figure 3.16. Average gain in Composite test scores, by group

Next, we examine score declines by mode of learning group, subject, and grade level (Figure 3.17). For all three mode of learning groups, score declines (in standard deviation units) were largest for math relative to the other subjects, with a few exceptions. The gap between Virtual and other groups was clearly largest for math, but the gap decreased with grade level. For English, reading, and science, the gap between Virtual and the other learning modes was smaller. For all subjects, grade levels, and mode of learning groups, scores declined relative to the 2019 group.

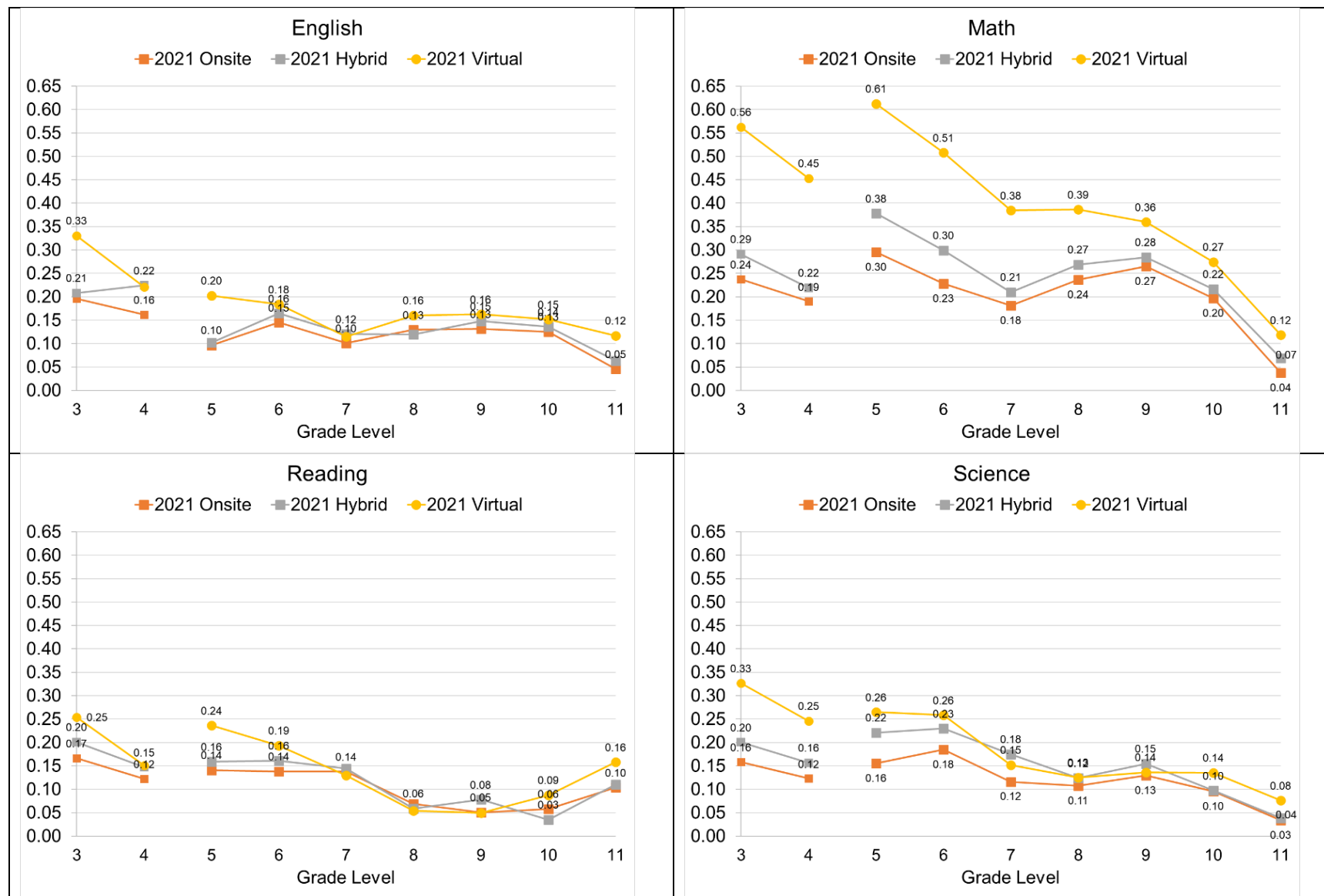


Figure 3.18. Decline in average test scores (in standard deviation units) by group, subject, and grade level

3.4.3 Results by Student Group

Mode of learning differences were also examined for different student groups. In this section, we highlight Composite score results for two groups that had mode of learning differences that varied from the patterns observed for the total group. Figure 3.19 shows the decline in average Composite scores for African American and White students, by mode of learning group and grade level. The mode of learning differences were substantially smaller for African American students, especially for higher grade levels. In particular, the Virtual gap for grades 8–11 was much smaller for African American students relative to White students. For African American students, the differences between Onsite and Hybrid were particularly small.

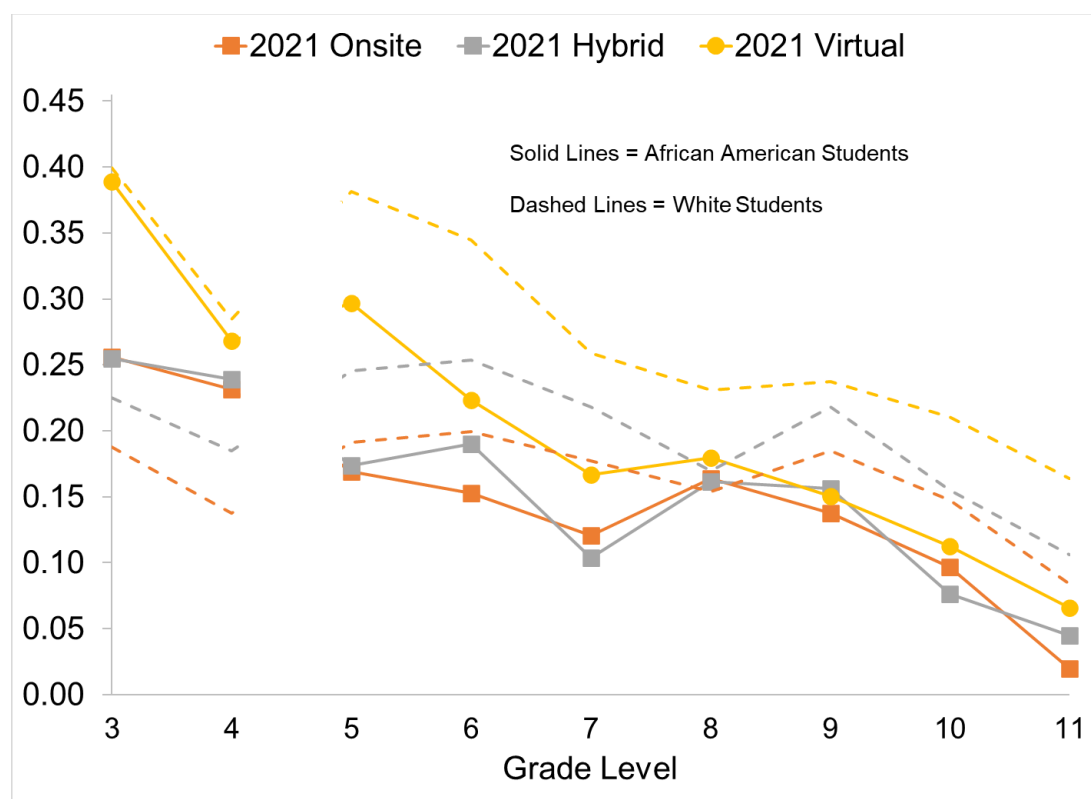


Figure 3.19. Decline in average Composite test scores (in standard deviation units) by mode of learning, for African American and White students

While the mode of learning differences for smaller African American students, this does not imply that Hybrid and Virtual modes of learning were more effective for African American students relative to others. To illustrate this point, we examine the average Composite score gain for African American students relative to students in other racial/ethnic groups (Table 3.6). Results are averaged across grades 5–10. We see that the baseline (2019) gain was 2.50 for African American students and 3.61 for other

racial/ethnic groups, for a difference of 1.11 score points. For each 2021 mode of learning group, the differences were less severe. However, in all cases, students in other racial/ethnic groups still had higher average gain scores than African American students, suggesting that Hybrid and Virtual modes were not more effective for African American students.

Table 3.6. Composite gain scores for African American and students in other racial/ethnic groups, by mode of learning group

Group	African American	Other Racial/Ethnic Groups	Difference
2019	2.50	3.61	-1.11
2021 Onsite	1.40	2.31	-0.91
2021 Hybrid	1.62	2.11	-0.49
2021 Virtual	1.16	1.63	-0.47

Figure 3.20 shows the decline in average Composite scores for students with disabilities and students without disabilities, by mode of learning group and grade level. For all grade levels, the differences across mode of learning groups was smaller for students with disabilities (solid lines). For 8th and 10th grade, the differences were extremely small. On the other hand, the mode of learning differences for students without disabilities were large and consistent across grade levels, with the Onsite group scoring higher than the Hybrid group, and the Hybrid group scoring higher than the Virtual group.

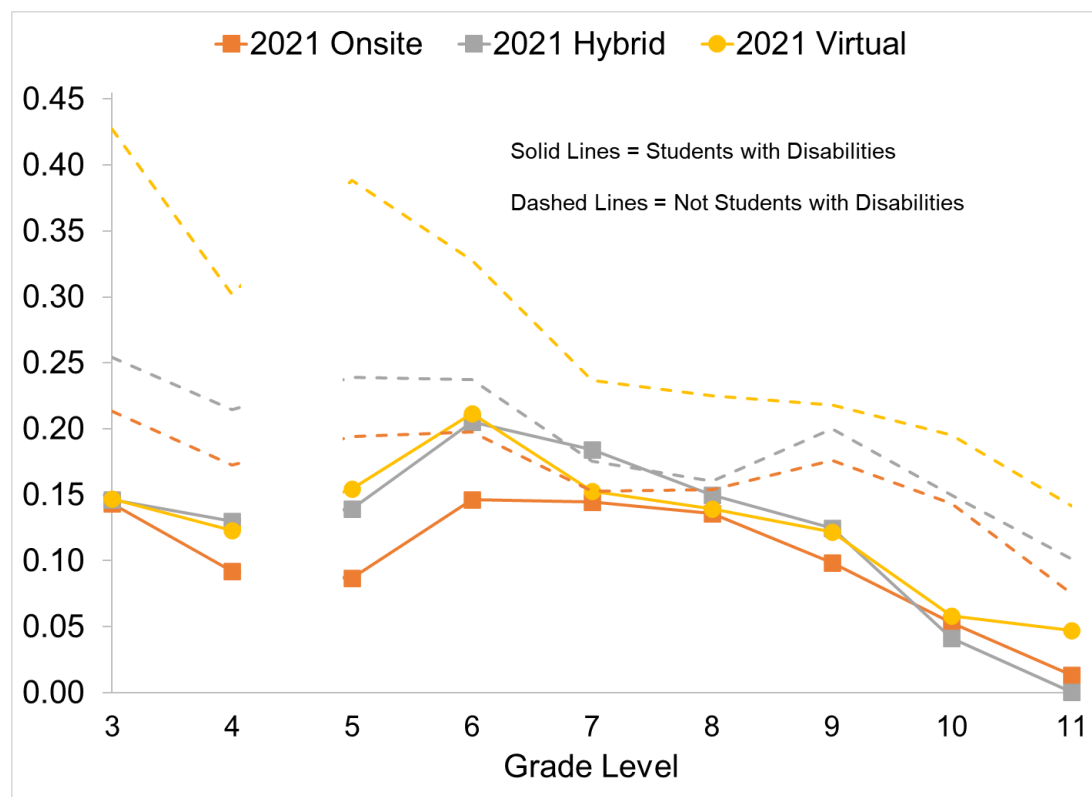


Figure 3.20. Decline in average Composite test scores (in standard deviation units) by mode of learning and disability status

3.4.4 Summary of Performance Differences by Mode of Learning

- Using data collected at four time points during the 2020–2021 school year, most mode of learning classifications were All Onsite (59%), All Hybrid (10%), or All Virtual (15%).
- African American and Asian students were less likely to learn onsite.
- Relative to 2019, ACT Aspire and ACT test scores declined in 2021 for all three mode of learning groups.
- Score declines were most severe for students who learned virtually.
- Score declines were less severe for students who learned onsite relative to those who learned in hybrid mode.
- The virtual gap was most severe for math and for lower grade levels.
- Mode of learning differences were small for reading.

- Mode of learning differences were relatively smaller for African American students and students with disabilities.

Chapter 4: Technical Characteristics of the Tests

This chapter discusses the technical characteristics of the ACT Aspire Summative Assessments, including score equating procedures and the analysis results for reliability and measurement error using data from the spring 2021 ACT Aspire administration in Arkansas. A description of the Arkansas students who participated in testing in spring 2021 is provided in [Chapter 2](#) of this technical report.

4.1 Test Equating

The ACT Aspire Summative Assessments system, testing students in English, mathematics, reading, science, and writing, is designed to measure student achievement and key areas of college and career readiness. Multiple ACT Aspire test forms have been developed with rigorous procedures and equated based on large, representative samples. Despite being constructed to follow the same content and statistical specifications, test forms may differ slightly in difficulty. The psychometric equating procedure is used to adjust for these differences in difficulty across forms so that scale scores reported to students have the same meaning regardless of the specific form administered. Through equating, statistical adjustments are made to maintain score interchangeability across test forms (see Holland & Dorans, 2006; Kolen & Brennan, 2014).

4.1.1 Equating Data Collection Design

For ACT Aspire, equating studies are conducted separately for online and paper testing modes. To ensure ACT Aspire scale scores are comparable regardless of testing mode, the paper base form was linked to the online base form through a comparability study conducted in spring 2013 using a random groups design with equipercentile equating. Similarly, in each equating study, a random groups design is typically used. In this case, test forms, including the anchor and new forms, are interspersed at the student level within a testing room so that the forms are distributed evenly and administered to randomly equivalent groups of students. Under a successful implementation of this design, observed differences in test performance across forms can be attributed to differences in form difficulty and equating methods can be applied to adjust for these differences. For each equating study, spiraling occurs separately for paper and online test forms. Across subjects and grades, the sample size of students taking each form ranges from 4,000 to 40,000 or more.

Under certain special situations, ACT Aspire also uses a common-item nonequivalent group design to equate test forms. For example, in the situation where test specifications are modified to better align to content standards, anchor forms may be revised, and a common-item equating design is implemented to collect student data across administrations. A revised anchor form is first equated to its original version using a common-item nonequivalent groups design, and then the new forms are equated to the revised anchor form using a random equivalent groups design.

Before equating is performed, collected equating data are checked for spiraling appropriateness and answer key correctness, and irregular student testing behaviors are reviewed. For each ACT Aspire equating study, depending on the equating design, the test level statistics of the anchor forms are examined and found to be similar across years. Based on data from the 2018 ACT Aspire administration in Arkansas, a review of the raw and scale score distributions of the Arkansas test-taking groups indicates that they are comparable to those of the equating samples.

4.1.2 Equating Methodology

For ACT Aspire, scores on parallel test forms are equated and placed on the vertical score scale using an equipercentile equating methodology (e.g., Kolen and Brennan 2014). In equipercentile equating, scores on different test forms are considered equivalent if they have the same percentile rank in a given group of students. Equipercentile equating is applied to the raw number-of-points scores for each subject test separately. The equipercentile equating results are subsequently smoothed using an analytic method described in Kolen (1984) to establish a smooth curve of the equivalents which are then rounded to integers. The conversion tables that result from this process are used to transform raw scores on the new forms to scale scores.

In addition to the scale scores of subject tests, ACT Aspire reports other scores including the Composite score, the ELA score, the STEM score, and the reporting category scores. These reported scores are not equated directly. The Composite, ELA, and STEM scores are a rounded arithmetic average of the scale scores from the contributing tests. Within each grade level, they are comparable across test forms because the scores used to compute them have been equated. The reporting category scores are calculated based on the number of earned points and are not equated across forms.

4.2 Reliability and Measurement Error

For any educational assessment program, an examinee might obtain different scores when tested with parallel forms on different administration occasions. The variation in

scores may reflect random sources of measurement error such as test anxiety, motivation, and distraction. In this context, reliability refers to the consistency of scores across replications of a measurement process. As an index of reliability and precision of measurement, coefficient alpha and scale score reliability were computed using Arkansas Aspire student test data. Under classical test theory (CTT), coefficient alpha is used to estimate the reliability of the test scores and indicates the internal consistency of items on a test. Similarly, scale score reliability is a concept which relates error-score variance and observed-score variance.

The standard error of measurement (SEM) is closely related to test reliability. SEM summarizes the amount of error or inconsistency in test scores. In interpreting an examinee's score, it is helpful to know the SEM of the test score. See Chapter 11 of the technical manual for more discussion on reliability and SEM.

Classification consistency indices quantify the reliability of categorizing examinees into mastery or achievement levels, with respect to specific performance standards. Several model-based approaches have been developed for estimating classification consistency for a single test administration because repeated testing data are seldom available. See Chapter 11 of the technical manual for discussion on the approach for analyzing classification consistency. For this report, classification consistency is examined with respect to the ACT Readiness Benchmarks and ACT Readiness Levels, which are introduced in [Section 3.1.2](#) of this report and described in more detail in Chapter 9 of the technical manual.

For Arkansas, after receiving the operational data from the spring 2021 Aspire administration, ACT psychometricians reviewed the data and conducted data cleaning, as needed, for reliability and measurement error analyses. Analysis results are presented as follows.

4.2.1 Raw Score Reliability and SEM for Subject Tests

For the English, mathematics, reading, and science tests, raw score reliability estimates (coefficient alpha) based on the Arkansas student test data from the ACT Aspire standard online forms administered in 2021 are presented in Table 4.1. Coefficient alpha for the writing test was not estimated as it is composed of a single prompt.

The reliability estimates were computed for the overall student population, the gender groups (female and male), and the major student groups identified for accountability reporting (African American, Hispanic, White, Economically Disadvantaged, Students with Disabilities, and English Language Learners).

For the overall student population, the observed raw score reliability estimates were larger than 0.80 across all subjects and grades. The observed small sizes of the ranges indicated that the raw score reliability estimates were comparable across forms. In comparison to the national reliability analysis results for online forms, the estimates based on the Arkansas student test data were all in the observed range of national estimates except for the 9th and 10th grade mathematics tests and 9th grade reading tests. See Chapter 11 of the technical manual for the ranges of the estimates of raw score reliability and SEM based on the national test data. As shown in Table 4.1, the reliability estimates calculated using the data for each student group were within acceptable levels given the test length and the homogeneity of students in each group under analysis.

Table 4.1. Ranges of Raw Score Reliability (Coefficient Alpha) Across Forms, by Subject and Grade Level

Subject	Grade level	Number of items	All	Female	Male
English	3	31	0.83	0.83	0.83
	4	31	0.86	0.86	0.86
	5	31	0.83	0.82	0.83
	6	35	0.86	0.85	0.85
	7	35	0.85	0.85	0.85
	8	35	0.85	0.85	0.84
	9	50	0.90	0.90	0.90
	10	50	0.91	0.91	0.91
Mathematics	3	30	0.86	0.85	0.87
	4	30	0.82	0.80	0.84
	5	30	0.81	0.79	0.83
	6	36	0.85	0.84	0.87
	7	36	0.85	0.85	0.86
	8	42	0.87	0.86	0.88
	9	42	0.86	0.84	0.87
	10	42	0.88	0.86	0.89
Reading	3	24	0.87	0.86	0.87
	4	24	0.86	0.85	0.87
	5	24	0.86	0.86	0.87
	6	24	0.84	0.83	0.85
	7	24	0.86	0.85	0.87
	8	24	0.83	0.82	0.84
	9	24	0.81	0.79	0.83
	10	24	0.82	0.80	0.84
Science	3	32	0.90	0.89	0.90
	4	32	0.87	0.87	0.88
	5	32	0.89	0.88	0.90
	6	36	0.90	0.90	0.91
	7	36	0.90	0.89	0.90
	8	36	0.90	0.89	0.91
	9	36	0.89	0.88	0.90
	10	36	0.91	0.90	0.91

Table 4.1. (continued)

Subject	Grade level	Number of items	African American	Hispanic	White
English	3	31	0.75	0.78	0.84
	4	31	0.80	0.83	0.86
	5	31	0.76	0.80	0.83
	6	35	0.80	0.83	0.86
	7	35	0.80	0.83	0.85
	8	35	0.78	0.82	0.85
	9	50	0.87	0.88	0.90
	10	50	0.88	0.90	0.91
Mathematics	3	30	0.79	0.84	0.86
	4	30	0.69	0.78	0.81
	5	30	0.66	0.77	0.81
	6	36	0.76	0.82	0.85
	7	36	0.76	0.81	0.85
	8	42	0.76	0.83	0.87
	9	42	0.76	0.82	0.85
	10	42	0.80	0.85	0.87
Reading	3	24	0.82	0.84	0.87
	4	24	0.80	0.83	0.87
	5	24	0.82	0.84	0.86
	6	24	0.79	0.81	0.84
	7	24	0.83	0.84	0.86
	8	24	0.77	0.80	0.83
	9	24	0.75	0.79	0.80
	10	24	0.77	0.81	0.82
Science	3	32	0.84	0.87	0.89
	4	32	0.83	0.84	0.86
	5	32	0.85	0.85	0.88
	6	36	0.87	0.88	0.90
	7	36	0.85	0.87	0.89
	8	36	0.86	0.87	0.90
	9	36	0.84	0.87	0.89
	10	36	0.85	0.88	0.90

Table 4.1. (continued)

Subject	Grade level	Number of items	Econ. Dis.	SWD	ELL
English	3	31	0.80	0.77	0.64
	4	31	0.84	0.81	0.72
	5	31	0.81	0.78	0.63
	6	35	0.84	0.80	0.69
	7	35	0.83	0.80	0.66
	8	35	0.82	0.79	0.61
	9	50	0.89	0.87	0.77
	10	50	0.90	0.89	0.82
Mathematics	3	30	0.84	0.86	0.79
	4	30	0.78	0.82	0.68
	5	30	0.77	0.81	0.62
	6	36	0.82	0.84	0.71
	7	36	0.82	0.84	0.66
	8	42	0.84	0.87	0.69
	9	42	0.82	0.86	0.66
	10	42	0.85	0.87	0.75
Reading	3	24	0.85	0.83	0.76
	4	24	0.84	0.82	0.72
	5	24	0.85	0.84	0.74
	6	24	0.82	0.81	0.70
	7	24	0.85	0.84	0.74
	8	24	0.81	0.80	0.67
	9	24	0.79	0.79	0.68
	10	24	0.81	0.81	0.71
Science	3	32	0.88	0.90	0.81
	4	32	0.86	0.88	0.79
	5	32	0.88	0.90	0.79
	6	36	0.89	0.91	0.82
	7	36	0.88	0.90	0.78
	8	36	0.89	0.91	0.78
	9	36	0.88	0.89	0.72
	10	36	0.89	0.90	0.78

Note. Econ. Dis. = Economically Disadvantaged; SWD = Students with Disabilities; ELL = English Language Learner.

4.2.2 Scale Score Reliability and SEM for Subject Tests, Composite, ELA, and STEM Scores

Table 4.2 presents the ranges of scale score reliability and SEM for English, mathematics, reading, and science by grade level across forms. The observed scale score reliability estimates based on the Arkansas student test data were similar across forms. Across subjects and grades, the estimates were larger than 0.80 except for some forms for the 5th grade English, 4th and 5th grade mathematics, and 8th grade reading tests for which the values of the observed scale score reliability were slightly smaller than 0.8. Within a subject and a grade level, the magnitude of the scale score reliability was comparable to the national estimates presented in the technical manual. The values of SEM were marginally larger than those from the national results and tended to become larger as the grade level progressed, similar to the finding in the national results.

Table 4.3 contains the scale score reliability and SEM for the ACT Aspire Composite, ELA, and STEM scores by grade level. The scale score reliability estimates for these combined scores were fairly high with values of 0.90 or larger and were in line with the national analysis results. See Chapter 11 of the technical manual for the national scale score reliability analysis results.

Table 4.2. Ranges of Scale Score Reliability and SEM Across Forms, by Subject and Grade Level

Subject	Grade level	Number of items	Reliability	SEM
English	3	31	0.81	2.75
	4	31	0.82	2.80
	5	31	0.78	3.22
	6	35	0.82	3.49
	7	35	0.81	3.78
	8	35	0.80	4.35
	9	50	0.88	3.53
	10	50	0.90	3.51
Mathematics	3	30	0.80	1.83
	4	30	0.77	2.09
	5	30	0.73	2.49
	6	36	0.80	2.39
	7	36	0.81	3.02
	8	42	0.85	3.00
	9	42	0.83	3.24
	10	42	0.85	3.22
Reading	3	24	0.82	2.27
	4	24	0.84	2.51
	5	24	0.85	2.57
	6	24	0.83	2.80
	7	24	0.83	2.81
	8	24	0.75	3.91
	9	24	0.81	3.43
	10	24	0.83	3.38
Science	3	32	0.89	2.33
	4	32	0.85	2.56
	5	32	0.86	2.42
	6	36	0.88	2.37
	7	36	0.87	2.64
	8	36	0.88	2.74
	9	36	0.86	3.11
	10	36	0.88	3.05

Table 4.3. Scale Score Reliability and SEM for the ACT Aspire Composite, ELA, and STEM Scores, by Grade Level

Grade level	Composite		ELA		STEM	
	Reliability	SEM	Reliability	SEM	Reliability	SEM
3	0.95	1.19	0.92	1.53	0.93	1.41
4	0.95	1.29	0.93	1.48	0.90	1.65
5	0.95	1.28	0.92	1.62	0.89	1.73
6	0.94	1.48	0.91	1.93	0.91	1.80
7	0.95	1.57	0.93	1.76	0.92	2.07
8	0.95	1.75	0.90	2.14	0.93	2.09
9	0.96	1.57	0.93	1.93	0.92	2.25
10	0.96	1.67	0.94	1.83	0.94	2.14

4.2.3 Raw Score Reliability and SEM for Reporting Category Scores

Within each ACT Aspire subject test, items are grouped by reporting categories, representing different components and topics covered by the test. The reporting categories correspond to the strands used to group the ACT College and Career Readiness Standards and ACT Aspire Performance Level Descriptors. See Chapter 3 of the technical manual for more details on the ACT Aspire reporting categories.

ACT Aspire score reports include reporting category scores, showing how students performed on different areas within each test. The reporting category scores are intended to foster a deeper understanding of strengths and weaknesses within a subject area, and to make connections to descriptions of the knowledge and skills students are likely to have. For each reporting category, the percentage and number of points students earn out of the total number of points possible are calculated and reported. The number of points possible for each reporting category may vary across forms, as indicated by the number of items in Tables 4.4 through 4.7.

The raw score reliability (coefficient alpha) and SEM for English, mathematics, reading, and science reporting category scores are provided in Tables 4.4, 4.5, 4.6, and 4.7, respectively. The observed raw score reliability estimates and SEM using the Arkansas student test data were comparable to those obtained from the national results. Within a subject and a grade, the ranges of the reliability and SEM estimated from the Arkansas student test data were quite similar to those estimated from the national data. See

Chapter 11 of the technical manual for the raw score reliability and SEM for reporting category scores based on the national data.

For both the Arkansas and national results, the reliability of reporting category scores is low for some of the reporting categories with very few items. It is important for test users to understand that reporting category scores are not intended for high-stakes decisions. Instead, the reporting category scores can help indicate which areas of the test students found most difficult and can be used as one source of evidence for identifying students' relative strengths and weaknesses.

Table 4.4. Ranges of Number of Items and Raw Score Reliability and SEM Across Forms for English Reporting Category Scores, by Grade Level

Grade level	Reporting category	Number of items	Reliability	SEM
3	POW	14	0.64	1.70
	COE	17	0.76	1.76
4	POW	9	0.65	1.35
	KLA	3	0.40	0.77
	COE	19	0.79	1.87
5	POW	9	0.72	1.29
	KLA	3	0.42	0.71
	COE	19	0.67	1.93
6	POW	11	0.67	1.39
	KLA	3	0.35	0.72
	COE	21	0.77	1.97
7	POW	10	0.67	1.36
	KLA	5	0.39	0.98
	COE	20	0.77	1.94
8	POW	11	0.70	1.45
	KLA	4	0.49	0.85
	COE	20	0.74	1.90
9	POW	13	0.74	1.53
	KLA	7	0.70	1.02
	COE	30	0.83	2.40
10	POW	13	0.77	1.51
	KLA	7	0.73	0.98
	COE	30	0.85	2.36

Note. POW = production of writing; KLA = knowledge of language; COE = conventions of standard English.

Table 4.5. Ranges of Number of Items and Raw Score Reliability and SEM Across Forms for Mathematics Reporting Category Scores, by Grade Level

Grade level	Reporting category	Number of items	Reliability	SEM
3	GLP	21	0.82	2.08
	NBT	3	0.61	0.63
	NF	4	0.37	0.87
	OA	6	0.66	1.01
	G	3	0.31	0.77
	MD	4	0.37	0.81
	IES	9	0.64	1.74
	JE	3	0.61	1.47
	MODELING	19	0.78	2.32
4	GLP	20	0.72	1.98
	NBT	5	0.58	0.93
	NF	5	0.37	0.99
	OA	3	0.03	0.74
	G	3	0.31	0.65
	MD	3	0.28	0.73
	IES	10	0.66	1.66
	JE	3	0.56	1.21
	MODELING	14	0.58	1.76
5	GLP	21	0.77	2.06
	NBT	5	0.44	1.03
	NF	5	0.54	0.86
	OA	3	0.31	0.79
	G	4	0.36	0.90
	MD	3	0.36	0.68
	IES	9	0.49	1.41
	JE	3	0.51	1.07
	MODELING	14	0.58	1.72
6	GLP	24	0.80	2.08
	NS	5	0.46	0.97
	EE	6	0.55	1.01
	RP	4	0.40	0.83
	G	4	0.55	0.80
	S	4	0.30	0.83
	IES	12	0.65	1.71
	JE	3	0.56	1.08
	MODELING	17	0.74	1.83

Table 4.5. (continued)

Grade level	Reporting category	Number of items	Reliability	SEM
7	GLP	24	0.79	2.21
	NS	4	0.38	0.84
	EE	6	0.58	1.00
	RP	5	0.34	0.97
	G	4	0.25	0.94
	S	4	0.50	0.84
	IES	12	0.65	1.65
	JE	3	0.59	1.18
	MODELING	17	0.70	1.77
8	GLP	28	0.81	2.48
	NS	3	0.14	0.77
	EE	7	0.57	1.17
	F	6	0.59	1.04
	G	7	0.49	1.12
	S	4	0.35	0.91
	IES	14	0.71	1.65
	JE	3	0.58	1.14
	MODELING	14	0.67	1.66
9	GLP	28	0.77	2.31
	N	3	0.34	0.71
	A	7	0.49	1.06
	F	6	0.39	1.07
	G	7	0.34	1.21
	S	4	0.39	0.89
	IES	14	0.73	1.80
	JE	3	0.62	1.06
	MODELING	21	0.78	1.96
10	GLP	28	0.81	2.30
	N	3	0.37	0.69
	A	7	0.59	1.07
	F	6	0.49	1.08
	G	7	0.38	1.20
	S	4	0.40	0.89
	IES	14	0.75	1.81
	JE	3	0.64	1.08
	MODELING	21	0.81	1.95

Note. GLP = grade level progress; NBT = numbers & operations in base 10; NF = numbers & operations—fractions; OA = operations & algebraic thinking; G = geometry; MD = measurement & data; IES = integrating essential skills; JE = justification & explanation; NS = the number system; EE = expressions & equations; RP = ratios & proportional relationships; S = Statistics & probability; F = functions; N = number & quantity; A = algebra.

Table 4.6. Ranges of Number of Items and Raw Score Reliability and SEM Across Forms for Reading Reporting Category Scores, by Grade Level

Grade level	Reporting category	Number of items	Reliability	SEM
3	KID	15	0.80	1.90
	CAS	7	0.67	1.15
	IOK	2	0.25	0.88
	TC	8	0.75	1.60
4	KID	14	0.79	1.77
	CAS	8	0.64	1.20
	IOK	2	0.35	1.19
	TC	11	0.75	1.92
5	KID	14	0.76	1.74
	CAS	8	0.73	1.15
	IOK	2	0.29	1.15
	TC	9	0.72	1.71
6	KID	15	0.75	1.91
	CAS	7	0.57	1.13
	IOK	2	0.41	1.05
	TC	13	0.72	2.05
7	KID	15	0.77	1.86
	CAS	7	0.70	1.11
	IOK	2	0.37	1.02
	TC	9	0.70	1.76
8	KID	14	0.77	1.70
	CAS	8	0.61	1.22
	IOK	2	0.20	1.51
	TC	8	0.64	1.88

Table 4.6. (continued)

Grade level	Reporting category	Number of items	Reliability	SEM
9	KID	14	0.74	1.94
	CAS	7	0.39	1.22
	IOK	3	0.38	1.38
	TC	12	0.67	1.88
10	KID	14	0.75	1.92
	CAS	7	0.45	1.20
	IOK	3	0.41	1.42
	TC	12	0.70	1.90

Note. KID = key ideas and details; CAS = craft and structure; IOK = integration of knowledge and ideas.

Table 4.7. Ranges of Number of Items and Raw Score Reliability and SEM Across Forms for Science Reporting Category Scores, by Grade Level

Grade level	Reporting Category	Number of items	Reliability	SEM
3	IOD	18	0.83	1.92
	SIN	7	0.61	1.28
	EMI	7	0.68	1.41
4	IOD	17	0.79	1.95
	SIN	7	0.55	1.31
	EMI	8	0.66	1.36
5	IOD	16	0.80	1.81
	SIN	8	0.67	1.28
	EMI	8	0.64	1.31
6	IOD	20	0.84	1.95
	SIN	7	0.66	1.26
	EMI	9	0.70	1.34
7	IOD	19	0.81	1.92
	SIN	7	0.61	1.37
	EMI	10	0.74	1.43
8	IOD	17	0.82	1.70
	SIN	7	0.65	1.19
	EMI	12	0.75	1.71
9	IOD	15	0.79	1.77
	SIN	9	0.64	1.38
	EMI	12	0.75	1.56
10	IOD	15	0.81	1.76
	SIN	9	0.67	1.38
	EMI	12	0.78	1.54

Note. IOD = interpretation of data; SIN = scientific investigation; EMI = evaluation of models, inferences, and experimental results.

4.2.4 Classification Consistency for the ACT Readiness Benchmarks and the ACT Readiness Levels

For the ACT Aspire English, mathematics, reading, and science tests, both the ACT Readiness Benchmarks and the ACT Readiness Levels are used for estimating classification consistency rates. Table 4.8 presents the classification consistency rates across forms by subject and grade level. The small sizes of the ranges indicated that the observed classification consistency rates were comparable across forms, and close to the lower bound of the range of classification consistency rates estimated from the national test data.

For ACT Aspire ELA and STEM scores, ACT Readiness Benchmarks are used for computing the classification consistency rates. The results of classification consistency analyses for ACT Aspire ELA and STEM scores using Arkansas data are presented in Table 4.9. The observed classification consistency rates for ELA and STEM were fairly high and close to those obtained from the national test results. See Chapter 11 of the technical manual for the national classification consistency analysis results.

Table 4.8. Classification Consistency Rates for ACT Readiness Benchmarks and ACT Readiness Levels Across Forms, by Subject and Grade Level

Subject	Grade level	Number of items	ACT Readiness Benchmark	ACT Readiness Levels
English	3	31	0.79	0.56
	4	31	0.82	0.59
	5	31	0.80	0.57
	6	35	0.82	0.59
	7	35	0.84	0.61
	8	35	0.82	0.58
	9	50	0.85	0.63
	10	50	0.86	0.65
Mathematics	3	30	0.80	0.56
	4	30	0.79	0.57
	5	30	0.79	0.57
	6	36	0.80	0.59
	7	36	0.81	0.56
	8	42	0.84	0.60
	9	42	0.86	0.64
	10	42	0.88	0.69
Reading	3	24	0.87	0.65
	4	24	0.84	0.61
	5	24	0.86	0.64
	6	24	0.85	0.62
	7	24	0.84	0.62
	8	24	0.79	0.51
	9	24	0.85	0.61
	10	24	0.85	0.64
Science	3	32	0.87	0.67
	4	32	0.83	0.60
	5	32	0.85	0.63
	6	36	0.86	0.64
	7	36	0.84	0.62
	8	36	0.85	0.64
	9	36	0.86	0.64
	10	36	0.87	0.67

Table 4.9. Classification Consistency Rates for ACT Readiness Benchmarks for ACT Aspire ELA and STEM Scores, by Grade Level

Grade level	ELA	STEM
3	0.90	0.93
4	0.89	0.92
5	0.89	0.95
6	0.88	0.95
7	0.89	0.93
8	0.87	0.94
9	0.89	0.95
10	0.90	0.95

Chapter 5: Validity Evidence

According to the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association, & National Council on Measurement in Education, 2014), “validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests” (p. 11). Validation is the process of justifying particular interpretations and uses, and it may involve logical, empirical, or theoretical components.

In this chapter, evidence of the validity of ACT Aspire scores for the proposed uses (described in [Section 1.3](#)) is presented. Validity evidence is often organized into the following six areas, as described by The Standards (AERA et al., 2014):

1. content
2. cognitive processes
3. internal structure
4. relationships with conceptually related constructs
5. relationships with criteria
6. consequences of tests

This chapter includes evidence related to content, internal structure, relationships with conceptually related constructs, and relationships with criteria.

5.1 Content-Oriented Evidence

ACT Aspire scores are intended to provide inferences about students’ knowledge and skills in English, mathematics, reading, science, and writing. Therefore, one aspect of validation for ACT Aspire is gathering content evidence for the foundational interpretation that ACT Aspire scores are indicative of academic achievement in English, mathematics, reading, science, and writing. Content evidence is important for all uses of ACT Aspire scores and is arguably the most important class of evidence for supporting the use of ACT Aspire scores for measuring progress toward meeting the Arkansas academic standards. Chapter 12 of the technical manual includes a summary of content-oriented validity evidence.

One component of content-oriented evidence is *alignment*, which generally refers to the degree that assessments and learning expectations are in agreement. With appropriate alignment, the state’s assessment program and framework for learning expectations work in tandem to support the educational system. Next, a summary is provided of an

independent study of the alignment of ACT Aspire's test content and performance level descriptors to the Arkansas state standards.

5.1.1 Alignment Study

The following description of the alignment study is taken from the *Executive Summary* and *Summary* sections of the report titled "[Alignment of the ACT® Aspire® Summative Assessments to the Arkansas Content Standards, Final Report, July 17, 2019](#)" by Susan Davis-Becker. We encourage interested readers to refer to the report for more details on the study's methodology and results.

ACT and the Arkansas Department of Education contracted ACS Ventures, LLC (ACS) to design, lead, and report on an independent alignment study to evaluate the alignment of the ACT Aspire Performance Level Descriptors (PLDs) and test content to the Arkansas academic standards for grades 3–8 and high school in English Language Arts (English, reading, & writing), mathematics, and science.

The study design was created around three key questions:

1. What level of cognitive processing is expected for students at each grade level for each standard or expectation?
2. How do the ACT Aspire PLDs reflect the knowledge and skills defined within the Arkansas standards AND demonstrate a level of cognitive complexity consistent with the level deemed appropriate/necessary for that standard(s)?
3. How does the ACT Aspire test content measure the knowledge and skills defined within the Arkansas standards AND demonstrate a level of cognitive complexity consistent with the level deemed appropriate/necessary for that standard(s)?

Educators from across the state of Arkansas convened during a four-day review to provide their expert judgments to answer these questions for the Aspire assessments in each subject area and at each grade level. The key findings from the subsequent analysis of their ratings are summarized below for each research question.

What level of cognitive processing is expected for students at each grade level for each standard or expectation?

Across all three subject areas, the panels found that the standards would likely require students to demonstrate a range of depth of knowledge (DOK) levels at each grade level. For ELA and mathematics, these targets were largely centered on levels 1–3 with a few specific standards targeting DOK level 4. For Science, these targets were slightly higher as the performance expectations are written to be encompassing of multiple expectations.

How do the ACT Aspire PLDs reflect the knowledge and skills defined within the Arkansas standards AND demonstrate a level of cognitive complexity consistent with the level deemed appropriate/necessary for that standard(s)?

The PLDs represent the full range of knowledge and skills that could be assessed from this domain on a form of the ACT Aspire but are likely larger and more all-encompassing than any one test form could be. Therefore, although aligning the PLDs provides a different perspective (compared to the test content), these results should be interpreted with caution as some of the elements within these descriptors are less specific than how test items are operationalized. That being said, the panels largely found alignment of the of the PLD elements to the Arkansas content standards.

- In ELA, the panels largely found the PLD elements to align and had similar expectations in terms of DOK. In reading specifically, several of the lower grade panels found a number of elements that were identified as either aligning to a lower grade level or not aligning.
- In mathematics, most panels found the majority of PLD elements to align to the Arkansas standards (and/or mathematical practices) and had similar expectations in terms of DOK. At the higher grade levels, a substantial percentage of the PLD elements were found to best align to lower grade level standards.
- In science, the panels largely found alignment to the science and engineering practices and the crosscutting concepts and had somewhat similar expectations in terms of DOK as the DOK expectations for each grade level were on the higher end of the scale whereas the PLD elements were largely in the middle of the scale. They also concluded that the specific elements could likely be aligned to any/all of the disciplinary core ideas depending on the context in which they were written.

How does the ACT Aspire test content measure the knowledge and skills defined within the Arkansas standards AND demonstrate a level of cognitive complexity consistent with the level deemed appropriate/necessary for that standard(s)?

Each panel reviewed three forms of the test to have a sample of the knowledge and skills that could be tested as well as the possible differences that may exist across forms. The results presented across subjects and grades identified a number of similarities but also some differences across the forms.

- In ELA, the panels largely found the test items to align and had similar expectations in terms of DOK. In reading, several of the lower grade panels found a number of items (testing specific concepts) that were identified as either aligning to a lower grade level or not aligning. In addition, there was some alignment to the anchor standards indicating the skills being measured were part of the overall goals for student learning but not specific to the focal grade.
- In mathematics, most panels found the majority of test items to align to the Standards (and/or mathematical practices) and had similar expectations in terms of DOK. At the higher grade levels, a substantial percentage of the items were found to best align to lower grade level standards.
- In science, the panels largely found alignment to the science and engineering practices and the crosscutting concepts and had somewhat similar expectations in terms of DOK as the DOK expectations for each grade level were on the higher end of the scale whereas the items were largely in the middle of the scale. They also concluded that the items did not directly align to the disciplinary core ideas based on how each was operationalized through the performance expectations.

5.2 Relationships With Conceptually Related Constructs

Often the intended interpretations of test scores imply that the scores should be correlated with conceptually related constructs (AERA et al., 2014). This section provides correlations of ACT Aspire test scores with two measures of conceptually related constructs: Partnership for Assessment of Readiness for College and Careers (PARCC) test scores and ACT test scores.

5.2.1 Correlations of ACT Aspire and PARCC Scores

PARCC test scores and ACT Aspire test scores measure related constructs, as both are standardized tests of the skills needed for college and career readiness (Pearson, 2017). For students in Arkansas, ACT Aspire and PARCC test scores were correlated

for PARCC tests taken in spring 2015 and ACT Aspire tests taken in spring 2016. PARCC ELA tests taken in grades 3–9 were paired with ACT Aspire ELA scores for grades 4–10, PARCC mathematics tests taken in grades 3–8 were paired with ACT Aspire mathematics tests taken in grades 4–9, and PARCC science tests taken in 5th and 7th grades were paired with ACT Aspire science tests taken in 6th and 8th grades.

Table 5.1 presents the samples sizes, test score means and standard deviations, and cross-grade correlations of the ACT Aspire and PARCC test scores. Across all subjects and grades, the correlations ranged from 0.71 to 0.84, and the disattenuated correlations ranged from 0.81 to 0.91. Correlations were similar across grades. On average, disattenuated correlations were highest in ELA (0.89), followed by mathematics (0.86) and science (0.85). Test scores were highly correlated, which is evidence that indirectly supports the use of ACT Aspire for measuring progress toward meeting the Arkansas academic standards and for determining if Arkansas students are on target for college and career readiness.

Table 5.1. Correlations of ACT Aspire Scores with PARCC Scores

Subject	Grade level pair	N	ACT Aspire		PARCC		r	r_{dis}
			Mean	SD	Mean	SD		
ELA	3–4	32,557	419.2	5.0	729.3	36.9	0.80	0.88
	4–5	32,891	421.5	5.5	736.2	31.4	0.81	0.90
	5–6	32,691	423.7	6.1	735.5	29.9	0.80	0.88
	6–7	32,810	423.3	6.3	736.7	29.0	0.82	0.89
	7–8	33,295	424.6	6.4	735.7	33.5	0.83	0.91
	8–9	32,725	425.1	7.0	733.9	34.0	0.83	0.89
	9–10	31,708	426.8	7.2	739.1	33.7	0.84	0.89
Mathematics	3–4	32,995	416.0	4.2	734.2	30.5	0.75	0.86
	4–5	33,137	417.9	5.2	729.9	27.9	0.76	0.89
	5–6	33,103	420.9	5.8	729.2	27.6	0.71	0.81
	6–7	32,950	420.5	6.8	730.5	27.1	0.76	0.86
	7–8	33,201	423.0	7.5	730.1	24.5	0.80	0.90
	8–9	26,816	421.9	6.5	721.3	29.7	0.74	0.82
Science	5–6	33,219	422.0	7.3	208.5	42.7	0.76	0.86
	7–8	33,554	423.6	7.8	181.0	47.1	0.74	0.83

Note. r = Pearson correlation; r_{dis} = disattenuated Pearson correlation.

In addition to the total group correlations (Table 5.1), correlations were examined for the student groups used for accountability reporting. For each group, the average sample size (across grades) is given in Table 5.2.

Table 5.2. PARCC/ACT Aspire Correlation Sample Sizes for Student Groups, Averaged Across Grades, for ELA, Mathematics, and Science

Group	ELA	Mathematics	Science
Total	32,668	32,034	33,387
African American	6,350	6,307	6,556
Hispanic	3,880	3,939	4,026
White	20,378	19,798	20,797
Economically Disadvantaged	19,677	19,863	20,333
Students with Disabilities	3,255	3,449	3,383
English Language Learners	2,519	2,682	2,666

For the total group and each student group, PARCC/ACT Aspire cross-grade correlations are presented in Table 5.3. The simple and disattenuated correlations were averaged across grades. For each group, weights were applied to make the group's distribution of lower-grade scores similar to the total group's distribution.

Table 5.3. PARCC/ACT Aspire Cross-Grade Correlations, Averaged Across Grades, for ELA, Mathematics, and Science

Group	ELA		Mathematics		Science	
	r	r_{dis}	r	r_{dis}	r	r_{dis}
Total	0.82	0.89	0.75	0.86	0.75	0.85
African American	0.80	0.87	0.72	0.82	0.72	0.81
Hispanic	0.82	0.89	0.73	0.83	0.74	0.83
White	0.82	0.89	0.76	0.86	0.74	0.84
Economically Disadvantaged	0.81	0.88	0.74	0.84	0.73	0.82
Students with Disabilities	0.81	0.88	0.74	0.84	0.68	0.77
English Language Learners	0.79	0.86	0.72	0.82	0.72	0.81

Note. r = Pearson correlation; r_{dis} = disattenuated Pearson correlation.

For ELA, there is very little variation across groups, with disattenuated correlations ranging from 0.86 (English Language Learners) to 0.89 (Hispanic and White groups and Total). Similarly, for math, disattenuated correlations ranged from 0.82 (African American and English Language Learners) to 0.86 (White group and Total). For science, the correlation for the Students with Disabilities group was 0.77, which was less than the Total group correlation (0.85). All of the group-specific correlations were within 0.10 of the total group's correlation. These findings suggest that ACT Aspire and

PARCC scores are highly correlated for all student groups with some minor variation in correlations across student groups.

5.2.2 Correlations of ACT Aspire and 11th Grade ACT Test Scores

ACT Aspire and the ACT both intend to measure the knowledge and skills most important for success in college and careers (ACT, 2020b). ACT Aspire is intended for earlier grades but is aligned to the same college and career readiness standards as the ACT and tests the same subjects as the ACT. If ACT Aspire and the ACT measure related constructs, high correlations would be expected between the two sets of test scores. Because the ACT is a commonly used measure of college readiness, high correlations of ACT Aspire scores and 11th grade ACT scores directly support the use of ACT Aspire scores for determining if Arkansas students are on target for college and career readiness.

We examined correlations of ACT Aspire and 11th grade ACT scores collected through spring 2021 for students from Arkansas. Because the ACT Aspire was first administered in spring 2016, only correlations for ACT Aspire grades 6–10 (6th grade students in spring 2016 took the ACT as 11th grade students in spring 2021) could be estimated.

Table 5.4 presents the sample sizes, test score means and standard deviations, and correlations of ACT Aspire and ACT scores for students from Arkansas. Note that the 6th grade sample includes one cohort of students (those who took Aspire in 6th grade in 2016 and the ACT in 11th grade in 2021). Similarly, the 7th, 8th, and 9th grade samples include two, three, and four cohorts of students, respectively. Because ACT Aspire testing was cancelled in 2020, rather than five, the 10th grade sample also includes four cohorts of students. The sample sizes are higher for higher grade levels because more cohorts of students are included.

Correlations generally increased with the ACT Aspire grade level, which is expected because the time between Aspire and ACT testing is shorter for higher grade levels. For example, the correlation of Composite scores was 0.81 for 6th grade, 0.83 for 7th grade, 0.85 for 8th grade, 0.87 for 9th grade, and 0.88 for 10th grade. The correlations ranged from 0.67 for 6th grade math and science to 0.88 for 10th grade Composite. The correlations suggest that ACT Aspire scores are strong predictors of ACT scores.

The disattenuated correlations of 10th grade ACT Aspire scores and 11th grade ACT scores ranged from 0.84 for reading to 0.91 for English and Composite. Because the correlation coefficients are very large, the findings indicate that ACT Aspire and the ACT

measure similar constructs. The disattenuated correlations also increased with ACT Aspire grade.

Table 5.4. Arkansas Correlations of ACT Aspire Scores with 11th Grade ACT Scores

Subject	Grade level	N	ACT Aspire		ACT		r	r_{dis}
			Mean	SD	Mean	SD		
English	6	19,186	425.5	7.7	17.6	6.2	0.71	0.82
	7	40,531	427.8	8.2	17.9	6.2	0.73	0.84
	8	63,461	428.2	8.8	17.9	6.3	0.76	0.87
	9	89,717	428.6	9.7	18.0	6.3	0.81	0.90
	10	94,361	430.6	10.2	18.2	6.4	0.82	0.91
Mathematics	6	19,200	421.5	5.7	17.8	4.4	0.67	0.78
	7	40,525	421.7	7.0	18.0	4.5	0.74	0.85
	8	63,581	424.4	7.7	18.0	4.5	0.78	0.88
	9	90,169	424.8	7.8	18.1	4.5	0.79	0.88
	10	94,532	426.1	8.3	18.3	4.5	0.80	0.89
Reading	6	18,914	419.7	6.8	18.5	6.3	0.68	0.80
	7	39,892	420.4	6.3	18.7	6.2	0.68	0.80
	8	62,941	423.2	7.1	18.8	6.3	0.69	0.82
	9	88,790	422.4	7.6	18.8	6.2	0.72	0.84
	10	93,078	423.6	7.8	19.0	6.2	0.72	0.84
Science	6	19,184	422.9	7.0	18.8	5.2	0.67	0.78
	7	40,118	423.0	7.4	18.9	5.2	0.69	0.81
	8	63,124	424.8	7.7	18.9	5.2	0.72	0.83
	9	88,858	425.9	8.3	18.9	5.2	0.74	0.85
	10	92,958	427.0	8.8	19.0	5.2	0.75	0.86
Composite	6	19,105	422.6	6.0	18.2	4.9	0.81	0.84
	7	40,538	423.4	6.5	18.4	5.0	0.83	0.86
	8	63,599	425.3	7.0	18.4	5.0	0.85	0.88
	9	89,999	425.6	7.5	18.5	5.0	0.87	0.90
	10	94,805	427.1	8.0	18.7	5.0	0.88	0.91
ELA	6	445	426.8	5.4	20.0	5.4	0.76	0.83
	7	641	427.9	6.0	20.1	5.4	0.78	0.85
	8	968	429.3	5.8	20.6	5.5	0.76	0.83
	9	1,335	430.6	6.3	21.1	5.8	0.82	0.89
	10	1,259	432.9	6.1	22.1	5.8	0.82	0.88
STEM	6	19,220	422.5	5.9	18.5	4.5	0.76	0.82
	7	40,455	422.6	6.7	18.7	4.5	0.79	0.85
	8	63,534	424.9	7.3	18.7	4.6	0.81	0.87
	9	89,799	425.6	7.6	18.7	4.5	0.83	0.89
	10	94,142	426.9	8.1	18.8	4.5	0.84	0.90

Note. r = Pearson correlation; r_{dis} = disattenuated Pearson correlation.

5.3 Relationships With Criteria

Intended uses of test scores imply that the scores should be predictive of criterion variables that are hypothetically related to the construct measured by the test. In this section, we examine how well ACT Aspire scores predict performance in high school courses, as well as performance on Advanced Placement (AP) exams.

5.3.1 Prediction of High School Course Grades

High school courses help students meet academic standards and prepare for college and careers. Thus, by measuring academic standards important for college and career readiness, ACT Aspire test scores should predict high school course grades—including standard, career-focused, AP, and dual-enrollment courses. By examining AP and dual-enrollment courses, college readiness is directly addressed. Similarly, by examining career-focused courses, career readiness is directly addressed. Students who are struggling in high school courses are candidates for extra academic support. ACT Aspire test scores, if predictive of performance in high school courses, can help to earlier identify students in need of support.

ACT Aspire scores from 8th, 9th, and 10th grade were linked to performance in high school courses, using data provided by the Arkansas Department of Education. ACT Aspire scores from spring 2016 and spring 2017 were linked to performance in the next year's courses (e.g., academic years 2016–2017 and 2017–2018). Analyses were conducted for 48 courses, including 19 standard courses, nine career-focused courses, 10 AP courses, and 10 dual-enrollment courses. For English and social science (social studies) courses, the ACT Aspire ELA score was used as the predictor. For mathematics and science courses, the ACT Aspire mathematics and science scores were used, respectively. For career-focused courses, the ACT Aspire Composite score was used. Analyses were also conducted by student group (Total group, African American, Hispanic, White, Economically Disadvantaged, Students with Disabilities, and English Language Learners).

Student grades were categorized as A, B, C, D, or F. For course grade data provided on a numeric scale (0–100), grades were coded as $A = \geq 90$, $B = 80\text{--}89$, $C = 70\text{--}79$, $D = 60\text{--}69$, and $F < 60$ or withdrawal from course. Three dichotomous grade outcomes were defined, representing different levels of success: A, B or higher, and C or higher. In this report, results for the B or higher criterion are summarized. Hierarchical logistic regression was used to relate the ACT Aspire test scores to the course success outcome. The model accommodates school-specific intercepts, which is important because grading standards vary across schools. ACT Aspire test scores were standardized (Mean = 0, SD = 1) for the population of Arkansas examinees at the grade

before the course was usually taken. Table 5.5 presents the sample sizes, logistic regression slope estimates, and accuracy rates for 27 of the 48 courses. Results for all 48 courses are available in a separate report (Allen, Radunzel, & Li, 2019).

Table 5.5. Logistic Regression Slopes and Accuracy Rates for Predicting B or Higher Grades in High School Courses

Course	Grade levels	Test score (Predictor)	N	Slope		INS accuracy Rate
				EST	SE	
English 9	8–9	ELA	61,575	1.47	0.01	0.75
English 10	9–10	ELA	59,460	1.40	0.01	0.74
English 11	10–11	ELA	39,232	1.16	0.02	0.71
Algebra I	8–9	Mathematics	46,635	1.44	0.02	0.69
Geometry	9–10	Mathematics	58,471	1.52	0.01	0.74
Algebra II	10–11	Mathematics	51,459	1.24	0.01	0.71
Physical Science	8–9	Science	60,569	1.37	0.01	0.74
Biology	9–10	Science	62,629	1.29	0.01	0.72
Chemistry	10–11	Science	40,715	1.19	0.02	0.71
Civics	8–9	ELA	58,940	1.24	0.01	0.74
World History	9–10	ELA	50,440	1.24	0.01	0.73
U.S. History	10–11	ELA	45,650	1.16	0.02	0.71
CF: Survey of Agriculture Systems	8–9	Composite	12,707	1.18	0.03	0.76
CF: Child Development	9–10	Composite	8,722	1.38	0.04	0.74
CF: Financial Literacy	10–11	Composite	5,481	1.41	0.05	0.77
AP English Lang. & Composition	10–11	ELA	14,874	1.60	0.04	0.76
AP Calculus AB	10–11	Mathematics	1,011	1.39	0.16	0.73
AP Biology	10–11	Science	3,197	1.45	0.08	0.74
AP Human Geography	8–9	ELA	1,566	1.87	0.12	0.69
AP World History	9–10	ELA	9,052	1.69	0.05	0.74
AP Psychology	10–11	ELA	1,662	1.79	0.13	0.78

Table 5.5. (Continued)

Course	Grade levels	Test score (Predictor)	N	Slope		INS accuracy Rate
DE: Oral Communication	10–11	ELA	1,739	1.07	0.09	0.69
DE: College Algebra	10–11	Mathematics	1,088	1.16	0.14	0.69
DE: Biology	9–10	Science	568	1.21	0.13	0.71
DE: Anatomy & Physiology	10–11	Science	330	1.18	0.20	0.70
DE: World History	9–10	ELA	735	1.47	0.13	0.74
DE: U.S. History	10–11	ELA	1,084	0.87	0.13	0.66

Note. EST = slope estimate; SE = standard error of slope estimate; INS = In Need of Support; AP = Advanced Placement; CF = career focused; DE = dual enrollment.

The logistic regression slopes represent the change in the log-odds of success, associated with a standard deviation increase in ACT Aspire test score. For example, for English 9, the slope estimate is 1.47. This means that the odds of earning a B or higher increase by a factor of 4.3 ($e^{1.47}$) for each standard deviation increase in ACT Aspire ELA score.

To evaluate how well ACT Aspire scores predict performance in high school courses, points of reference for the predictive strength of established measures are needed. Because ACT scores have been used to predict college success for decades with well-documented validity evidence (ACT, 2020b), we use results from ACT's College Readiness Benchmarks research as points of reference (Table 5.6) (Allen, 2013; Radunzel, Westrick, Bassiri, & Li, 2017). The logistic regression slope values presented in Table 5.6 were calculated using the same methods (hierarchical logistic regression) as used for this study. These slopes are referred to as *ACT reference slopes*.

Table 5.6. ACT Reference Slopes: Logistic Regression Slopes of ACT Scores for Predicting B or Higher Grades in College Courses

Subject	College course(s)	Slope
ELA	English composition	0.90
	Social science	1.23
Mathematics	College algebra	1.12
Science	Biology	1.15

Across the 27 courses presented in Table 5.5, the logistic regression slope ranged from 0.87 (for dual-enrollment U.S. History courses) to 1.87 (for AP Human Geography). For all courses but one, the logistic regression slopes exceeded each B or higher ACT reference slope. Thus, the relationship between ACT Aspire test scores and high school course performance is generally stronger than the relationship between ACT test scores and college course performance.

Using all 48 courses, the average logistic regression slopes for the B or higher criterion were compared across student groups (Table 5.7). For each group, weights were applied to make the group's distribution of test scores similar to the total group's distribution.² For all groups, the mean slope exceeded each ACT reference slope. Some variation was found across groups in the predictive strength of ACT Aspire test scores: Mean slopes were largest for the African American and White groups; slopes were smallest for the Students with Disabilities and English Language Learner groups. These results suggest that the predictive strength of ACT Aspire scores varies by group, but that ACT Aspire scores are strong predictors of success in high school courses for all student groups.

Table 5.7. Average Predictive Strength (Logistic Regression Slopes), by Student Group

Group	Number of courses	Average slope for predicting B or higher grades
Total	48	1.30
African American	36	1.31
Hispanic	36	1.20
White	48	1.30
Economically Disadvantaged	48	1.21
Students with Disabilities	26	0.98
English Language Learners	29	1.05

For each content area, course success rates can also be examined by ACT Aspire Readiness Levels (In Need of Support, Close, Ready, and Exceeding). B or higher success rates, averaged across core subject courses, are presented in Figure 5.2. As expected, success rates increased significantly with readiness level. Students at the *In Need of Support* level were the least likely to earn B or higher grades, with success rates ranging from 28% for mathematics courses to 43% for social studies courses.

² The weights correct for artificial differences across groups in mean slopes that can be attributed to the distribution of test scores.

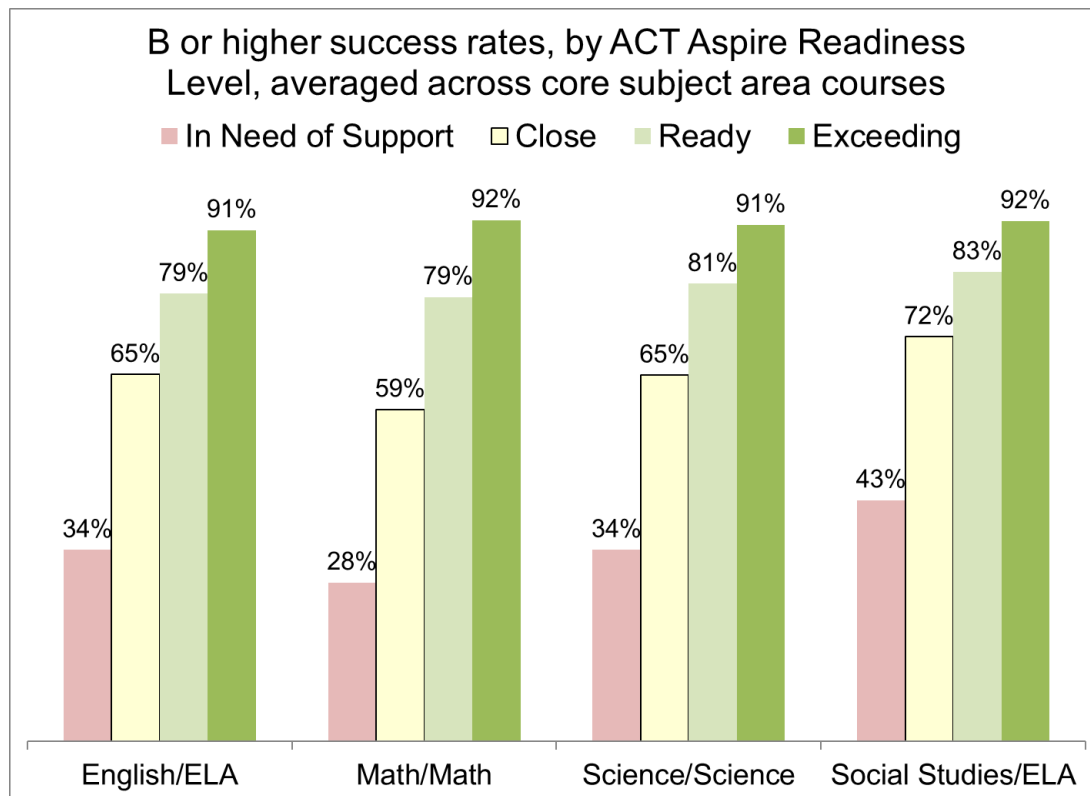


Figure 5.2. B or higher success rates, by ACT Aspire Readiness Level, averaged across core subject courses.

Comparing ACT Aspire Readiness Levels and course success outcomes enable calculations of accuracy rates and other measures of predictive strength. Table 5.8 shows the cross-tabulation of ACT Aspire Readiness Level and B or higher outcomes, averaged across courses. The table gives the percentage of students with each combination of readiness level and outcome, as well as margin percentages. Classifications of *In Need of Support* are considered accurate if a student is not successful in a course (i.e., earns a C or lower), and classifications of other readiness levels are considered accurate if a student is successful in a course (i.e., earns a B or higher) (see cells shaded in light green). Table 5.5 presents the *In Need of Support* accuracy rates for the 27 courses shown. Across all mathematics courses, the average accuracy rate for *In Need of Support* classifications was 71%. Similarly, the average accuracy rates were 74% for English courses, 72% for science courses, 73% for social studies courses, and 75% for career-focused courses.

Table 5.8. Comparing ACT Aspire Readiness Level and B or Higher Course Outcome, Averaged Over Courses

Course content area/ ACT Aspire test score	Readiness level	Course outcome		Total (%)
		B or higher (%)	C or lower (%)	
English/ELA	In Need of Support	12.9	25.4	38.3
	Close	13.4	7.2	20.7
	Ready	15.9	4.2	20.1
	Exceeding	18.9	2.0	20.9
	Total	61.2	38.8	100.0
Mathematics/ Mathematics	In Need of Support	13.2	31.4	44.6
	Close	14.7	10.7	25.4
	Ready	13.3	3.9	17.2
	Exceeding	11.7	1.1	12.8
	Total	52.9	47.1	100.0
Science/Science	In Need of Support	15.1	29.1	44.2
	Close	14.5	7.9	22.3
	Ready	16.4	3.9	20.3
	Exceeding	12.1	1.1	13.2
	Total	58.0	42.0	100.0
Social studies/ELA	In Need of Support	16.3	21.9	38.2
	Close	14.8	5.9	20.7
	Ready	16.8	3.4	20.2
	Exceeding	19.3	1.6	20.9
	Total	67.2	32.8	100.0
Career-focused/ Composite	Bronze or below	18.3	13.1	31.4
	Silver	29.3	5.1	34.4
	Gold	21.7	1.2	22.9
	Platinum	11.1	0.2	11.3
	Total	80.4	19.6	100.0

Note. For the career-focused courses, career readiness levels are based on the Progress Toward Career Readiness indicator (Allen, 2018) instead of the ACT Readiness Levels.

5.3.2 Prediction of High School GPA and Academic Rigor

In this section, we present additional research examining how well ACT Aspire scores predict performance in high school courses. Instead of considering individual high

school courses, we consider overall high school GPA a measure of academic rigor that accounts for the difficulty of courses taken.

ACT Aspire scores from 8th grade were linked to high school coursework and grades data reported by students when they registered for the 11th grade ACT test in 2021. For 30 different high school courses, students were asked to report the grade they earned in each course already taken, with five options (A, B, C, D, or F). For courses not yet taken, students were asked if they plan to take the course later in high school.

High school GPA (HSGPA) was calculated by averaging the grades reported by students. Only students who reported course grades in each core subject area (English, math, social studies, and natural science) were included in the analysis. On average, students reported grades in 14.4 of the 30 courses. When students register for the ACT test, they are also asked whether they have taken advanced placement, accelerated, or honors courses in English, mathematics, social studies, natural sciences, or foreign languages.

An index of course rigor was obtained using an item response theory (IRT) model known as the graded response model (Samejima, 1969). The model treats the 30 different courses as different items on a test and treats grades as the item scores. The model also treats the indicators for advanced coursework and student plans for taking upper-level STEM courses (chemistry, physics, advanced math, and calculus) as “items.” Under the IRT framework, the resulting estimates of student performance are calibrated across students with different (but sometimes overlapping) courses. The graded response model assumes that the course grade probability distribution is determined by course-specific discrimination and difficulty parameters, as well as a latent trait distribution that is assumed to have a mean of 0 and standard deviation of 1. The latent trait to represent student ability is often denoted using the theta symbol (θ) and we refer to this trait as the “rigor index.” Prior research has found that, relative to HSGPA, the rigor index has less skewness and higher correlations with college degree attainment (Allen & Mattern, 2019).

Table 5.9 presents the sample sizes, summary statistics for ACT Aspire test scores and HSGPA, and correlations of ACT Aspire test scores with HSGPA and the rigor index. Results are provided for each ACT Aspire subject, but note that HSGPA and the rigor index include all courses (not subject specific).

Table 5.9. Arkansas Correlations of 8th Grade ACT Aspire Scores with High School GPA and Academic Rigor

Subject	N	ACT Aspire		HSGPA		r HSGPA	r Rigor
		Mean	SD	Mean	SD		
English	11,826	430.8	8.6	3.30	0.60	0.48	0.54
Mathematics	11,783	427.1	7.7	3.30	0.60	0.50	0.58
Reading	11,611	424.9	6.6	3.29	0.60	0.47	0.52
Science	11,840	427.2	7.4	3.30	0.60	0.52	0.58
Composite	11,764	427.6	6.8	3.30	0.60	0.55	0.62
ELA	373	430.5	5.2	3.54	0.45	0.43	0.52
STEM	11,854	427.4	7.2	3.30	0.60	0.54	0.62

Note. r = Pearson correlation.

Correlations of ACT Aspire test scores with HSGPA ranged from 0.47 for Reading to 0.55 for Composite. Correlations of ACT Aspire test scores with the rigor index are even higher, ranging from 0.52 for Reading to 0.62 for Composite and STEM. The correlations suggest that ACT Aspire test scores from 8th grade are good predictors of overall performance in high school courses. When the rigor of the coursework is accounted for, the relationship between ACT Aspire scores and performance is even stronger.

5.3.3 Prediction of Success on AP Exams

If ACT Aspire test scores predict AP exam scores, there is additional evidence that ACT Aspire measures college readiness. Such evidence would directly support another use of ACT Aspire scores: to inform readiness for advanced high school coursework.

For this study, ACT Aspire scores from 8th, 9th, and 10th grade were linked to performance on AP exams, using data provided by the Arkansas Department of Education. Hierarchical logistic regression was used to relate ACT Aspire scores to success on the AP exam, which is defined as earning a 3 (“qualified for doing the work of an introductory-level college course”) or higher.

Analyses were conducted for different AP courses and student groups. There were 10 AP courses and six student groups (English Language Learners, Students with Disabilities, Economically Disadvantaged, African American, Hispanic, and White). For AP exams associated with English and social studies courses, the ACT Aspire ELA score was used as the predictor. For AP exams in mathematics and science, the ACT Aspire mathematics and science scores were used, respectively.

With 10 courses and seven groups (six student groups and total group), there are 70 possible conditions. We restricted the analysis to conditions with a sample size of at least 100, resulting in 45 conditions for analysis. ACT Aspire scores were standardized (Mean = 0, SD = 1) for the population of Arkansas examinees at the grade before the AP exam was usually taken.

Table 5.10 presents the sample sizes, logistic regression slope estimates, and accuracy rates for the 10 AP courses. The logistic regression slopes represent the change in the log-odds of success, which is associated with a standard deviation increase in ACT Aspire test score. For example, for AP Statistics, the slope estimate was 3.05. This means that the odds of earning a 3 or higher on the AP Statistics exam increase by a factor of 21.1 ($e^{3.05}$) for each standard deviation increase in ACT Aspire mathematics score. Across the 10 courses, the slope estimates range from 2.15 (Human Geography) to 3.83 (English Literature and Composition). Because the slopes are much larger than the ACT reference slopes (Table 5.6), the findings suggest that ACT Aspire scores are very strong predictors of success on AP exams.

Table 5.10. Logistic Regression Slopes and Accuracy Rates for Predicting Success (Score of 3 or Higher) on AP Exams

AP exam	Grade level pair	Test score (predictor)	N	Slope		Exceeding accuracy rate
				EST	SE	
English Lang. & Comp.	10–11	ELA	13,373	3.47	0.07	0.86
English Lit. & Comp.	10–11	ELA	1,923	3.83	0.22	0.85
Calculus AB	10–11	Mathematics	802	2.23	0.20	0.92
Statistics	10–11	Mathematics	813	3.05	0.25	0.94
Biology	10–11	Science	2,841	2.93	0.13	0.91
Chemistry	10–11	Science	1,025	2.90	0.24	0.90
Physics	10–11	Science	1,657	2.81	0.20	0.90
Human Geography	8–9	ELA	1,347	2.15	0.14	0.84
World History	9–10	ELA	7,873	2.39	0.07	0.84
Psychology	10–11	ELA	1,459	2.72	0.16	0.86

Note. EST = slope estimate; SE = standard error.

The probability of success on an AP test can be graphed by ACT Aspire score to show the predictive strength of ACT Aspire. For example, Figure 5.3 shows the probability of success on the AP English Language and Composition exam as a function of ACT Aspire ELA score. The probability of success is very low for ACT Aspire ELA scores

below 430 but then increases drastically for scores above 430. Students with ELA scores of 438 and higher have more than a 0.50 probability of succeeding on the AP exam, and students with ELA scores of 443 and higher have more than a 0.90 probability of succeeding on the AP exam.

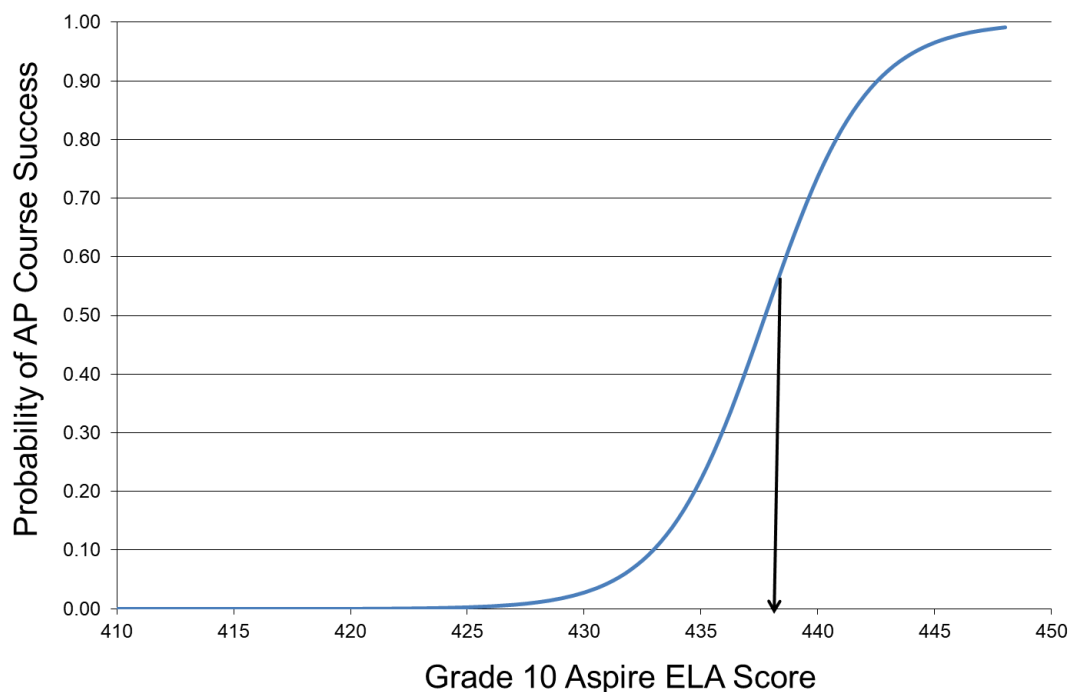


Figure 5.3. Probability of success (score of 3 or higher) on AP English Language and Composition Exam by ACT Aspire ELA score.

The average logistic regression slopes for AP exam success were compared across student groups (Table 5.11). Results are not provided for the Students with Disabilities group because no AP exams met the sample-size requirement of 100 or more students. For all other groups, the mean slopes far exceed all ACT reference slopes (Table 5.11). No significant differences in predictive strength across groups were found suggesting that, across student groups, ACT Aspire test scores are very strong predictors of success on AP exams.

Table 5.11. Predictive Strength (Logistic Regression Slopes) for AP Exam Success (Score of 3 or Higher), by Student Group

Group	Number of courses	Logistic regression slope estimates		
		Min	Mean	Max
Total	10	2.15	3.01	3.83
African American	6	2.19	3.33	5.14
Hispanic	7	2.34	2.89	3.45
White	10	2.04	2.95	3.82
Economically disadvantaged	10	2.20	3.01	3.94
English language learner	2	1.96	3.34	5.33

For each AP exam, course success rates can also be examined by ACT Aspire Readiness Levels (In Need of Support, Close, Ready, and Exceeding). Exam success rates are presented in Figure 5.4. As expected, success rates increased significantly with readiness level. Results are not shown for the *In Need of Support* level because very few students at that level took AP exams. Students scoring at the *Exceeding* level had reasonably high rates of success on AP exams, ranging from 25% for Physics to 66% for Biology and Psychology. Students at the *Close* level had very low rates of success on AP exams. The success rates for students scoring at the *Ready* level range from 3% (English Literature and Composition) to 27% (Psychology).

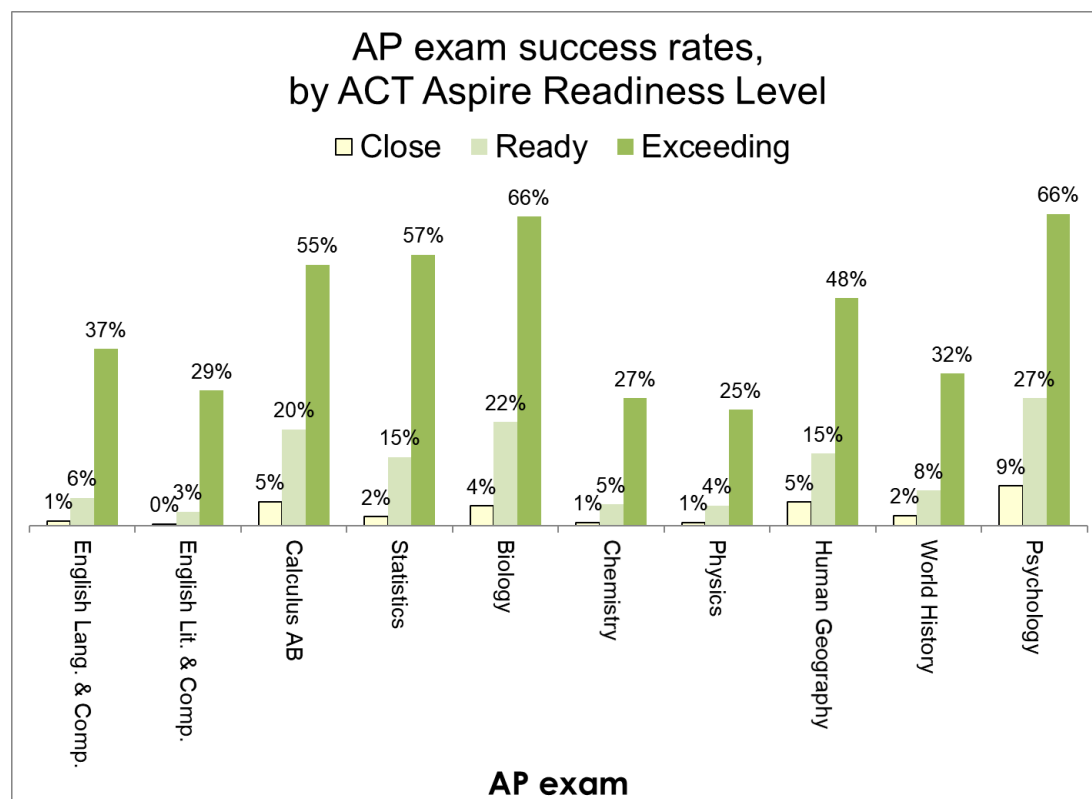


Figure 5.4. AP exam success rates, by ACT Aspire Readiness Level.

5.3.4 Prediction of First-Year College Outcomes

A recent study (Allen, 2021) examined the relationship between ACT Aspire scores and three types of first-year college outcomes: (a) taking remedial coursework, (b) grades in common first-year credit bearing courses, and (c) overall academic performance. If ACT Aspire scores are predictive of first-year college outcomes, it provides additional evidence that ACT Aspire Summative Assessments measure progress toward college and career readiness.

The study was based on two cohorts of students from Arkansas who enrolled in college at one of Arkansas' public colleges or universities. The first cohort took ACT Aspire as 10th graders in spring 2016, took the ACT as 11th graders in spring 2017, and enrolled in college in fall 2018. The second cohort took ACT Aspire as 9th graders in spring 2016 and 10th graders in spring 2017, took the ACT as 11th graders in spring 2018, and enrolled in college in fall 2019. The samples were weighted to be more similar to the Arkansas ACT Aspire tested population on gender, race/ethnicity, economic disadvantage status, disability status, and English learner status. The full study is documented in a report that is available [here](#). In this section, we summarize the study's main findings.

Finding 1: College remediation rates decreased with higher ACT Aspire scores.

The probability of taking remedial courses in English, mathematics, and reading drastically decreased as 9th and 10th grade ACT Aspire scores increased. For students scoring at the ACT Readiness Benchmarks, the probability of taking a remedial course was 0.32 for 10th grade English, 0.13 for 10th grade mathematics, and 0.06 for 10th grade reading. Additional analysis showed that the probability of taking remedial courses was much lower for students who met the corresponding ACT College Readiness Benchmark. This likely reflects the course placement policies adopted by Arkansas' public colleges and universities, which often use ACT scores and the ACT College Readiness Benchmarks when making placement decisions.

Finding 2: Grades in benchmark college courses increased with higher ACT Aspire scores.

Grades earned in benchmark college courses increased steadily with increasing ACT Aspire scores and readiness levels. The statistical relationship between ACT Aspire scores and probability of course success (defined as "B" or higher grade) was strongest for Biology/science and was weakest for English Composition I/English. ACT Aspire test scores from 10th grade were stronger predictors than those obtained from 9th grade, which is expected because the 10th grade test occurs in closer time proximity to the outcome.

Finding 3: ACT Aspire scores are predictive of overall first-year academic performance.

Table 5.12 provides the correlations of predictors with a measure of difficulty-adjusted first-year college GPA. The correlations represent the mean correlation observed across postsecondary institutions. In addition to the simple correlation (r), a correlation corrected for measurement error in the dependent variable is provided (r_c).

ACT Aspire Composite scores are predictive of overall first-year academic performance, with correlations comparable to those of ACT Composite scores and just below those of high school coursework and grades (measured by high school GPA and difficulty-adjusted high school GPA). This suggests that ACT Aspire Composite scores from 9th and 10th grade provide an early signal of overall readiness for college.

Table 5.12. Correlations with First-Year Academic Performance

Cohort	Predictor	r	r_c
2018	Aspire Composite 10th Grade	0.401	0.445
	ACT Composite 11th Grade	0.384	0.435
	High School GPA	0.428	0.498
	Difficulty-adjusted High School GPA	0.452	0.520
2019	Aspire Composite 9th Grade	0.390	0.432
	Aspire Composite 10th Grade	0.402	0.460
	ACT Composite 11th Grade	0.398	0.455
	High School GPA	0.433	0.512
	Difficulty-adjusted High School GPA	0.460	0.536

Finding 4: ACT Aspire scores are incrementally predictive of overall first-year academic performance, relative to high school grades and ACT Composite score.

In addition to the correlation analysis, regression models were fit to examine how well ACT Aspire scores help predict first-year academic performance, above difficulty-adjusted high school GPA and ACT Composite score. The first model, known as the Partial model, used difficulty-adjusted high school GPA and ACT Aspire Composite score as predictors and so tested the incremental prediction of ACT Aspire Composite score over difficulty-adjusted high school GPA. The second model, known as the Full model, used difficulty-adjusted high school GPA, ACT Composite score, and ACT Aspire Composite score as predictors and so tested the incremental prediction of ACT Aspire Composite score over difficulty-adjusted high school GPA and ACT Composite score.

Table 5.13 provides the standardized regression coefficients for predicting difficulty-adjusted first-year college GPA. The coefficients represent the mean coefficient observed across postsecondary institutions. In addition to the uncorrected coefficient (β), a coefficient corrected for measurement error in the dependent variable is provided (β_c). Table 5.13 also presents the multiple regression correlation coefficient (R), which measures each model's overall predictive strength. In addition, the increase in R after adding ACT Aspire Composite score(s) to the models is provided, denoted as ΔR .

ACT Aspire Composite scores were incrementally predictive of first-year academic performance, contributing to the model's prediction beyond both difficulty-adjusted high school GPA (Partial model) and the combination of difficulty-adjusted high school GPA and ACT Composite score (Full model). In the Full models, the predictive strength of ACT Aspire Composite scores was comparable to that of ACT Composite score. This

suggests that ACT Aspire test scores from 10th grade and perhaps earlier can be combined with ACT scores to form a stronger measure of college readiness based on standardized test scores, relative to using ACT scores alone, though the increase in predictive strength (R) after adding ACT Aspire Composite scores is very modest.

Table 5.13. Multiple Regression Predictors of First-Year Academic Performance

Cohort	Model	Predictor	β	β_c	$R, \Delta R$
2018	Partial	Aspire Composite 10th Grade	0.165	0.184	0.562, 0.025
		High School Academic Performance	0.336	0.387	
	Full	Aspire Composite 10th Grade	0.140	0.155	0.566, 0.003
		ACT Composite 11th Grade	0.107	0.121	
		High School Academic Performance	0.333	0.384	
2019	Partial	Aspire Composite 9th Grade	0.195	0.216	0.554, 0.018
		High School Academic Performance	0.362	0.421	
	Full	Aspire Composite 9th Grade	0.034	0.038	0.561, 0.002
		Aspire Composite 10th Grade	0.080	0.092	
		ACT Composite 11th Grade	0.126	0.144	
		High School Academic Performance	0.341	0.397	

5.4 Differential Item Functioning

According to the Standards for Educational and Psychological Testing (AERA et al., 2014), “analyses of the internal structure of a test can indicate the degree to which the relationships among test items and test components conform to the construct on which the proposed test score interpretations are based.” (p.16). Differential item functioning (DIF) analyses examine whether items on a test may function differently for identifiable groups of examinees. Results of DIF analyses can be used as one form of validity evidence based on the internal structure of the test.

An item is flagged for DIF when examinees from one group have a higher probability of responding correctly than examinees from another group with the same ability. The procedures used for the analysis of the 2021 Arkansas student test data include the Mantel-Haenszel common odds-ratio (MH) procedure and the standardized difference in proportion-correct (STD) procedure. Established guidelines were used to classify DIF for each item into one of three levels: A (insignificant DIF), B (slight to moderate DIF), or C (moderate to large DIF). Chapter 13 of the technical manual provides more information on the DIF classification rules.

The DIF analysis results presented in Tables 5.14 and 5.15 include the percentage of items that met DIF A, DIF B, and DIF C flagging criteria for the gender and race/ethnicity (White/African American, White/Hispanic) comparisons, respectively.

Most of the items reviewed using the Arkansas student test data were flagged at the A level, regardless of the comparisons. For gender comparisons, the percentages of DIF B level items were slightly larger than 10% for mathematics across all grades (except 6th and 7th), and for 6th grade reading as well as 5th grade science. For the White/African Americans comparisons, only mathematics at 6th, 8th, and 10th grade had percentages of DIF B level items larger than 10%. The percentages of DIF C level items were considered small for all the comparisons. In general, the DIF analysis results based upon the Arkansas student test data were comparable to those of the national analyses. It is noted that flagging an item does not mean the item is necessarily biased. Some items that are flagged and appear to favor one group over another might just be due to random fluctuations in samples. For ACT Aspire, items that are statistically flagged for the national analyses are further reviewed by content and measurement specialists to eliminate the potential concern of item bias. See Chapter 13 of the technical manual for the national DIF analysis results.

Table 5.14. Summary of Gender DIF Analysis, by Subject and Grade Level

Subject	Grade level	DIF classification (%) for Female vs. Male		
		A	B	C
English	3	100.0	0.0	0.0
	4	100.0	0.0	0.0
	5	100.0	0.0	0.0
	6	100.0	0.0	0.0
	7	97.1	2.9	0.0
	8	97.1	2.9	0.0
	9	98.0	2.0	0.0
	10	98.0	2.0	0.0
Mathematics	3	93.3	6.7	0.0
	4	80.0	20.0	0.0
	5	96.7	3.3	0.0
	6	88.9	5.6	5.6
	7	94.4	2.8	2.8
	8	95.2	4.8	0.0
	9	92.9	7.2	0.0
	10	88.1	11.9	0.0
Reading	3	91.7	8.3	0.0
	4	95.8	4.2	0.0
	5	91.7	8.3	0.0
	6	83.3	16.6	0.0
	7	95.8	4.2	0.0
	8	91.7	4.2	4.2
	9	91.7	4.2	4.2
	10	91.7	4.2	4.2
Science	3	96.9	3.1	0.0
	4	100.0	0.0	0.0
	5	96.9	3.1	0.0
	6	91.7	8.4	0.0
	7	97.2	2.8	0.0
	8	94.4	5.6	0.0
	9	94.4	5.6	0.0
	10	94.4	5.6	0.0

Note. Row percentages may not add up to 100% because of rounding.

Table 5.15. Summary of Ethnicity DIF Analysis, by Subject and Grade Level

Subject	Grade level	DIF classification (%) for African American vs. White			DIF classification (%) for Hispanic vs. White		
		A	B	C	A	B	C
English	3	96.8	3.2	0.0	93.5	6.5	0.0
	4	100.0	0.0	0.0	100.0	0.0	0.0
	5	100.0	0.0	0.0	96.8	3.2	0.0
	6	97.1	2.9	0.0	100.0	0.0	0.0
	7	100.0	0.0	0.0	100.0	0.0	0.0
	8	94.3	2.9	2.9	100.0	0.0	0.0
	9	100.0	0.0	0.0	100.0	0.0	0.0
	10	100.0	0.0	0.0	98.0	2.0	0.0
Mathematics	3	93.3	6.7	0.0	100.0	0.0	0.0
	4	96.7	0.0	3.3	100.0	0.0	0.0
	5	90.0	10.0	0.0	100.0	0.0	0.0
	6	97.2	2.8	0.0	100.0	0.0	0.0
	7	94.4	5.6	0.0	100.0	0.0	0.0
	8	97.6	2.4	0.0	100.0	0.0	0.0
	9	97.6	2.4	0.0	100.0	0.0	0.0
	10	97.6	2.4	0.0	100.0	0.0	0.0
Reading	3	100.0	0.0	0.0	100.0	0.0	0.0
	4	100.0	0.0	0.0	95.8	4.2	0.0
	5	100.0	0.0	0.0	91.7	8.3	0.0
	6	100.0	0.0	0.0	95.8	4.2	0.0
	7	95.8	0.0	4.2	91.7	4.2	4.2
	8	91.7	8.3	0.0	95.8	4.2	0.0
	9	91.7	8.3	0.0	91.7	8.3	0.0
	10	87.5	12.5	0.0	91.7	8.3	0.0
Science	3	100.0	0.0	0.0	100.0	0.0	0.0
	4	96.9	3.1	0.0	96.9	3.1	0.0
	5	96.9	3.1	0.0	96.9	3.1	0.0
	6	100.0	0.0	0.0	97.2	2.8	0.0
	7	97.2	2.8	0.0	100.0	0.0	0.0
	8	100.0	0.0	0.0	97.2	2.8	0.0
	9	100.0	0.0	0.0	97.2	2.8	0.0
	10	100.0	0.0	0.0	97.2	2.8	0.0

Note. Row percentages for a comparison may not add up to 100% because of rounding.

5.5 Depth of Knowledge Analysis

The cognitive complexity level of written passages and the cognitive demands of test items are important characteristics to consider when measuring a student's academic achievement. ACT Aspire assessments reflect the skills that students are expected to have to think, reason, and analyze at high levels of cognitive complexity. ACT Aspire items and tasks target different levels of cognitive complexity with most items targeted at upper levels.

Webb's Depth-of-Knowledge (DOK) system (2002) is widely used across the nation and in many educational contexts for understanding a test item's cognitive complexity. For ACT Aspire, ACT incorporates substantial training, discussion, and multiple inputs to achieve consistent implementation of cognitive complexity levels based on Webb's DOK language. The DOK levels are assigned to reflect the complexity of the cognitive process required, not the psychometric "difficulty" of the item. Unlike other DOK interpretations, ACT only assigns a DOK level 4 value to describe multiday, potentially collaborative classroom activities and assessments designed for learning purposes. By this definition, DOK assignments on any summative assessment including ACT Aspire are limited to values of 1 to 3.

ACT's DOK level 1 corresponds to Webb's level 1 where students are primarily actively using knowledge and skills with limited extended processing. ACT's DOK level 2 extends beyond level 1 and involves applying these cognitive processes to many situations, including real-world scenarios. Therefore, ACT's DOK level 2 aligns with Webb's DOK level 2 and some of Webb's DOK level 3. ACT's DOK level 3 involves situations where the student must apply high-level, strategic thinking skills to short- and long-term situations. Some of these situations are novel and some require generating something such as a graph, but all involve higher-level thinking skills. Given this interpretation, ACT's DOK level 3 aligns with Webb's DOK level 3 and DOK level 4.

Based on the spring 2021 Arkansas data, Table 5.16 contains the average percent correct by DOK level for each subject and grade. For mathematics and science at all grade levels, the observed average percent-correct decreased as the DOK level increased. The same trend was also observed for 3rd, 7th, and 8th grade English and for 7th, 9th, and 10th grade reading. As discussed previously, items with higher DOK are not necessarily more difficult than items with lower DOK. Generally, the relationship between item difficulty and DOK level, based on the Arkansas student test data, is strongest for mathematics and science, particularly for upper grade levels.

Table 5.16. Average Percent-Correct, by DOK Level, Subject, and Grade Level

Subject	Grade level	Depth of Knowledge (DOK) Level		
		1	2	3
English	3	56.6	55.8	41.5
	4	50.6	46.8	49.9
	5	53.6	52.8	58.2
	6	47.1	54.6	54.1
	7	60.5	47.3	51.9
	8	57.6	62.4	44.7
	9	53.9	54.8	58.0
	10	57.6	58.8	61.1
Mathematics	3	43.9	39.1	35.0
	4	53.8	38.7	35.1
	5	39.9	42.1	30.4
	6	44.8	45.4	28.9
	7	60.2	44.8	28.6
	8	40.3	40.3	28.4
	9	49.4	42.2	27.3
	10	54.9	46.1	28.8
Reading	3	37.9	52.0	44.0
	4	51.9	42.0	51.6
	5	53.6	52.3	39.2
	6	51.7	51.0	47.6
	7	60.0	56.6	55.5
	8	51.1	39.4	55.7
	9	48.4	53.3	45.0
	10	52.3	55.6	48.1
Science	3	54.2	49.0	49.8
	4	49.8	53.9	39.3
	5	59.6	60.9	45.9
	6	66.6	55.8	43.6
	7	65.6	56.1	40.5
	8	65.2	53.9	44.1
	9	67.5	49.4	33.0
	10	70.2	52.6	35.9

Chapter 6: Growth Summary

ACT Aspire Summative Assessments are designed to support interpretations of student growth through the following:

- vertical scaling of test scores across 3rd through 10th grade
- reporting of longitudinal progress charts with a student's current and prior year's scores in English, mathematics, reading, and science
- classification of a student's scores into ACT Readiness Levels, showing how a student scored relative to the ACT Readiness Benchmarks
- predicted paths, which predict a range of a student's ACT Aspire test scores over the next two years
- predicted 10th grade PreACT score range and predicted 11th grade ACT score range
- classification of student growth as low, average, or high on the basis of student growth percentiles (SGPs)

These features most directly support the use of ACT Aspire for determining if students are on target for college and career readiness and for assessing how well Arkansas schools and districts are preparing students for college and career (accountability). Chapter 14 of the technical manual provides more information on how ACT Aspire supports interpretations of student growth.

In this chapter, we summarize Arkansas-specific growth data, with comparisons to national norms. We also present an analysis showing how ACT Aspire Interim scores relate to the Summative scores and measures of student growth.

Note: Because ACT Aspire testing was cancelled for spring 2020, the growth data from spring 2018 to spring 2019 are the latest available one-year growth data for Arkansas. Therefore, the analyses presented in this chapter have not been updated for 2021.

6.1 Comparison of Mean Growth Scores to National Growth Norms

We compared the mean growth scores for Arkansas from spring 2018 to spring 2019 to the latest national norms. We focused on two types of growth scores: gain scores and SGPs. Gain scores support interpretations of absolute growth, and SGPs support comparisons of growth to norms established from a reference group.

Supported by the vertical scales developed for ACT Aspire, gain scores can be calculated as the arithmetic difference in scores from one year to the next. Positive mean gain scores are anticipated because students are expected to increase their knowledge and skills each year. SGPs represent the rank of a student's test score compared to the scores of students with the same prior year scores. ACT Aspire SGPs, ranging from 1 to 100, are available for students who test in consecutive years approximately one year apart. The primary use of ACT Aspire SGPs is to support growth comparisons across schools, subject areas, grade levels, instructional programs, and student groups. Such comparisons could lead to insights into what conditions are having greater effects on student learning.

SGPs are a normative measure of growth that must be interpreted with respect to a reference group of students. Each year, ACT updates the reference groups used to estimate the SGPs, using data from the most recent year of testing. Reference group samples are created for each subject and pair of adjacent grade levels, and each sample is designed to be representative of the ACT-tested population on race/ethnicity, school affiliation (public or nonpublic), and school percentage eligible for free or reduced lunch. The national growth norms used for this report are based on the samples used in the 2019 norming study, which include students from both public and nonpublic schools, including those from Arkansas.

For each grade level and subject, Table 6.1 provides the mean prior year (2018) score, mean current year (2019) score, mean gain score, and mean SGP for students from Arkansas. The mean gain score and mean SGP are also provided for the national reference samples.

Consistent with national norms, there is considerable variation across grades and subjects in mean gain scores for Arkansas. As expected, most of the mean gain scores are positive, showing that students typically increased their knowledge and skills after one year of schooling. For Arkansas, the few exceptions are for reading grade 8–9 (mean gain = -1.0) and ELA grade 6–7 (mean gain = -0.2). For both Arkansas and the national norms, mean gain scores tend to be larger for lower grade levels.

Table 6.1. 2019 Gain Score and SGP Means, by Subject and Grade Level

Subject	Grade level pair	Arkansas mean					National mean	
		N	Prior year score	Current year score	Gain	SGP	Gain	SGP
English	3–4	35,269	417.3	420.4	3.2	47.8	3.3	49.9
	4–5	36,256	420.4	423.2	2.7	46.7	3.1	49.8
	5–6	36,187	423.2	425.3	2.1	47.4	2.5	49.7
	6–7	33,750	425.4	427.8	2.4	47.9	2.6	49.6
	7–8	33,856	427.9	428.1	0.3	46.0	1.0	49.6
	8–9	33,462	428.1	428.5	0.4	46.4	1.0	49.5
	9–10	32,634	428.5	430.5	2.0	49.1	2.2	49.5
Mathematics	3–4	35,269	413.3	416.0	2.7	46.7	3.0	50.8
	4–5	36,251	416.0	417.5	1.6	46.1	2.1	50.3
	5–6	36,188	417.9	420.1	2.2	42.3	3.3	50.0
	6–7	33,748	420.7	421.8	1.1	53.7	0.5	49.7
	7–8	33,849	421.8	424.7	2.9	50.0	2.9	49.6
	8–9	33,456	424.5	424.8	0.3	44.1	1.3	49.5
	9–10	32,640	424.8	426.4	1.6	48.5	1.7	49.5
Reading	3–4	35,270	412.8	415.7	2.9	50.4	2.8	50.1
	4–5	36,256	415.5	417.7	2.2	49.0	2.3	50.0
	5–6	36,187	417.3	419.2	1.9	47.9	2.1	49.9
	6–7	33,762	419.3	420.0	0.8	48.0	0.8	49.9
	7–8	33,857	420.2	423.0	2.9	50.5	2.7	49.8
	8–9	33,454	422.9	421.9	-1.0	45.4	-0.3	49.7
	9–10	32,628	422.2	423.1	0.9	49.2	1.0	49.7
Science	3–4	35,260	415.3	417.8	2.6	46.6	2.9	49.9
	4–5	36,242	417.7	419.7	2.0	45.5	2.3	49.8
	5–6	36,175	419.6	421.1	1.5	47.8	1.7	49.9
	6–7	33,732	421.1	422.7	1.6	52.5	1.0	49.8
	7–8	33,850	422.6	424.5	1.9	48.1	2.1	49.7
	8–9	33,437	424.6	426.1	1.6	51.1	1.2	49.6
	9–10	32,632	425.8	427.3	1.5	50.2	1.4	49.6

Table 6.1. (continued)

Subject	Grade level pair	Arkansas Mean					National mean	
		N	Prior year score	Current year score	Gain	SGP	Gain	SGP
Composite	3–4	35,223	414.8	417.6	2.8	48.6	3.0	51.0
	4–5	36,211	417.5	419.6	2.1	47.7	2.4	51.3
	5–6	36,141	419.6	421.6	2.0	45.9	2.4	50.4
	6–7	33,686	421.7	423.2	1.5	51.9	1.3	50.2
	7–8	33,792	423.2	425.2	2.0	48.5	2.2	50.3
	8–9	33,354	425.2	425.5	0.3	45.6	0.8	50.1
	9–10	32,526	425.5	427.0	1.5	48.6	1.6	49.9
ELA	3–4	34,106	417.7	420.8	3.1	51.1	2.9	50.5
	4–5	35,540	420.4	422.7	2.3	48.4	2.5	50.4
	5–6	35,554	422.2	424.9	2.7	50.2	2.7	50.2
	6–7	33,233	424.8	424.6	-0.2	46.6	0.2	50.0
	7–8	33,593	424.4	426.2	1.7	50.7	1.6	50.0
	8–9	33,158	425.8	426.0	0.2	45.7	0.8	50.0
	9–10	32,215	426.4	427.4	1.1	47.6	1.4	49.9
STEM	3–4	35,247	414.5	417.2	2.6	46.6	2.9	50.5
	4–5	36,229	417.1	418.9	1.8	46.3	2.1	50.5
	5–6	36,162	419.0	420.9	1.9	45.1	2.4	50.2
	6–7	33,709	421.1	422.5	1.4	54.5	0.8	49.9
	7–8	33,822	422.5	424.9	2.4	48.7	2.5	49.9
	8–9	33,401	424.8	425.7	0.9	47.1	1.3	49.8
	9–10	32,592	425.6	427.1	1.5	49.1	1.6	49.7

Across 49 subject / grade level combinations, the mean gain for Arkansas was larger than the mean gain for the national reference sample in 12 cases and less than the mean gain for the national reference sample in 37 cases. It is important to keep in mind that the national reference samples are designed to be representative of the ACT-tested population, not the general population. The mean gain scores for Arkansas may compare more favorably to the general population.

By definition, the mean SGP for the national reference samples is close to 50 for all subject areas and grade levels (Table 6.1). For Arkansas, the mean SGP ranged from 42.3 (mathematics grade 5–6) to 54.5 (STEM grade 6–7). Across the 49 subject / grade level combinations, the mean SGP for Arkansas was 1.8 units lower than the national mean SGP. Similar to the gain score comparison, the mean SGP for Arkansas was larger than the mean SGP for the national reference sample in 12 cases and less than the mean SGP for the national reference sample in 37 cases.

6.2 ACT Readiness Level Transitions

As described in [Section 3.1.2](#), ACT Aspire’s Readiness Levels include four levels: **In Need of Support**, **Close**, **Ready**, and **Exceeding**. To better understand the percentage of students transitioning across readiness levels, we provide the relative frequency of each readiness level, conditional on the prior year’s readiness level in Appendix [Table A.3](#). The percentages in Table A.3 are based on the same data used for the gain score and SGP analyses (Arkansas students with scores in spring 2018 and spring 2019).

For an example of how to interpret the percentages in Table A.3, consider students in 4th grade who were at the *Ready* level in mathematics (see cells of Table A.3 in the red box). Most of the students (60%) remained at the *Ready* level in 5th grade, 6% improved to the *Exceeding* level, 31% dropped to the *Close* level, and 2% dropped to the *In Need of Support* level. Because the ACT Readiness Benchmarks for mathematics become more difficult for higher grade levels, negative transitions are more common for mathematics relative to the other subjects.

6.3 Relating ACT Aspire Interim Scores to Summative SGPs

ACT Aspire Interim Assessments can be administered at any time during the academic year to provide information about students’ progress toward end-of-year learning goals (ACT, 2020c). The interim tests can be thought of as shorter versions of the ACT Aspire Summative tests and cover the same subjects (English, math, reading, and science) and grades (3rd–10th) as the ACT Aspire Summative Assessments.

More information on the ACT Aspire Interim Assessments is provided in the [ACT Aspire Periodic Technical Manual](#).

Analyses were conducted to examine how performance on the interim tests throughout the latest academic year (2018–2019) relates to performance on the spring 2019 Summative tests and to growth from spring 2018 to 2019. Interim tests taken during the

academic year were grouped into time periods as: September/October, November through January, or February through April. The analysis only uses students who took at least one Interim test during each of the three time periods. Results are provided in Table 6.2 by subject and grade level.

Interim score averages are presented for each time period. Interim scores generally increased over the academic year. Consistent with Summative gain scores, Interim gains were generally higher in the lower grades.

Across all subjects and grade levels, the average correlation of Interim score and Summative score is 0.73 for September/October, 0.74 for November–January, and 0.74 for February–April. The average correlation of Interim score and growth from spring 2018 to spring 2019 (SGP) is 0.26 for September/October, 0.28 for November–January, and 0.30 for February–April. The positive correlations of Interim test scores with spring-to-spring growth is expected because students who perform well during the academic year should show more growth on Summative assessments.

The positive correlations of Interim performance and Summative SGPs suggest that students who perform better on the ACT Aspire Interim tests are more likely to demonstrate more growth on the ACT Aspire Summative tests. The evidence supports using Interim test scores and SGPs to indicate a student's progress over an academic year.

Table 6.3. Relationships of Interim and Summative Scores

Subject	Grade Level	N	Mean Interim scores			Correlations with Summative scores			Correlations with SGP		
			SEP-	NOV-	FEB-	SEP-	NOV-	FEB-	SEP-	NOV-	FEB-
English	3–4	5,282	158.9	159.2	159.9	0.73	0.73	0.75	0.32	0.34	0.36
	4–5	5,942	161.1	161.4	161.8	0.75	0.74	0.74	0.24	0.24	0.25
	5–6	5,813	162.9	163.3	163.4	0.76	0.76	0.78	0.32	0.34	0.35
	6–7	4,575	163.9	164.0	164.0	0.76	0.77	0.78	0.27	0.29	0.29
	7–8	4,514	161.9	161.8	162.0	0.76	0.74	0.76	0.33	0.33	0.36
	8–9	3,847	170.6	170.8	171.3	0.82	0.82	0.81	0.28	0.30	0.32
	9–10	3,619	171.7	171.9	172.3	0.83	0.82	0.81	0.29	0.28	0.31
Mathematics	3–4	6,155	159.0	159.0	160.3	0.61	0.69	0.72	0.24	0.31	0.34
	4–5	6,546	156.6	157.2	157.6	0.64	0.68	0.75	0.17	0.21	0.27
	5–6	6,599	157.5	158.0	158.8	0.65	0.63	0.72	0.26	0.25	0.30
	6–7	5,343	159.8	159.9	160.6	0.72	0.74	0.75	0.30	0.34	0.35
	7–8	4,939	161.3	161.7	162.0	0.76	0.79	0.77	0.21	0.24	0.23
	8–9	4,554	157.8	158.2	158.4	0.74	0.70	0.65	0.19	0.18	0.17
	9–10	4,759	159.0	159.4	159.3	0.78	0.74	0.70	0.17	0.17	0.17
Reading	3–4	6,024	158.2	158.5	159.0	0.72	0.72	0.71	0.33	0.35	0.37
	4–5	6,069	158.7	159.0	159.5	0.73	0.75	0.75	0.31	0.33	0.36
	5–6	5,734	158.1	158.5	158.7	0.73	0.72	0.73	0.26	0.28	0.30
	6–7	5,139	159.2	159.4	159.4	0.73	0.73	0.73	0.28	0.31	0.33
	7–8	4,994	158.6	158.6	159.0	0.70	0.70	0.73	0.32	0.32	0.34
	8–9	4,335	157.4	157.5	157.5	0.71	0.72	0.67	0.29	0.30	0.30
	9–10	4,219	157.9	158.1	158.1	0.70	0.72	0.67	0.25	0.26	0.29
Science	3–4	6,692	161.9	162.4	163.2	0.73	0.76	0.76	0.23	0.27	0.30
	4–5	7,154	163.9	164.5	165.2	0.71	0.74	0.74	0.21	0.24	0.27
	5–6	6,672	162.4	162.8	162.9	0.75	0.77	0.76	0.24	0.27	0.29
	6–7	6,132	163.7	164.3	164.5	0.74	0.77	0.76	0.24	0.28	0.30
	7–8	5,506	160.0	160.1	160.7	0.75	0.74	0.76	0.29	0.30	0.32
	8–9	5,221	161.3	161.0	161.5	0.75	0.74	0.76	0.21	0.24	0.26
	9–10	5,261	161.8	161.6	161.9	0.76	0.75	0.75	0.21	0.25	0.26

References

- ACT. (2020a). *ACT Aspire Summative technical manual* (August 2020 Version 2020.1). Iowa City, IA: Author.
- ACT. (2020b). *ACT technical manual*. (August 2020 Version 2020.1) Iowa City, IA: Author.
- ACT. (2020c). *ACT Aspire Periodic technical manual* (August 2020 Version 2020.1). Iowa City, IA: Author.
- Allen, J. (2013). *Updating the ACT College Readiness Benchmarks*. ACT Research Report 2013-6. Iowa City, IA: ACT, Inc.
- Allen, J. (2018). *Understanding updates to the ACT Aspire Progress Toward Career Readiness Indicator*. Iowa City, IA: ACT, Inc.
- Allen, J. (2021). *A longitudinal study of ACT Aspire performance and first-year college outcomes in Arkansas*. ACT Working Paper 2021-03. Iowa City, IA: ACT, Inc.
- Allen, J., Radunzel, J., & Li, J. (2019). *Relating ACT Aspire scores to performance in high school courses and other measures of college and career readiness*. ACT Research Report. Iowa City, IA: ACT, Inc.
- Allen, J. & Mattern, K. (2019). Examination of indices of high school performance based on the graded response model. *Educational Measurement: Issues and Practice*, 38(2), 41–52.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Austin, P.C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46(3), 399–424.
- Davis-Becker, S. (2019). Alignment of the ACT® Aspire® Summative Assessments to the Arkansas Content Standards, Final Report, July 17, 2019.
- Holland, P.W. & Dorans, N.J. (2006). Linking and equating. In R.L. Brennan (Ed.), *Educational measurement* (4th ed.) (pp. 187–220). Westport, CT: American Council on Education/Praeger Publishers.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.

Kolen, M.J. (1984). Effectiveness of analytic smoothing in equipercentile equating. *Journal of Educational and Behavioral Statistics*, 9(1), 25–44.

Kolen, M.J. & Brennan, R.L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). New York, NY: Springer-Verlag.

Pearson. (2017). *PARCC: Final technical report for 2016 administration*. Retrieved 3/20/2019 from <https://www.isbe.net/Documents/PARCC%202016%20Tech%20Report.pdf>.

Radunzel, J., Westrick, P., Bassiri, D., and Li, D. (2017). *Development and validation of a preliminary ELA readiness benchmark based on the ACT ELA score*. (ACT Research Report No. 2017-9). Iowa City, IA: ACT, Inc.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, 34, 100–114.

Webb, N. L. (2002, March 28th). Depth-of-knowledge levels for four content areas. Unpublished manuscript. Retrieved from: <https://apps.web.maine.gov/doe/sites/maine.gov.doe/files/inline-files/dok.pdf>.

Appendix

Table A.1. ACT Aspire Summative Technical Manual Chapters

Chapter Number and Title	Content
1. General description of ACT Aspire assessments and standards	Primary uses and claims of ACT Aspire, content frameworks, and standards alignment.
2. Test development	Assessment design and test development processes.
3. Assessment specifications	Each subject test's framework, reporting categories, scoring rubrics (for writing), item types / tasks, and test blueprints.
4. Item and task scoring	Procedures for scoring multiple-choice, technology-enhanced, constructed-response items; performance scoring quality control; and automated scoring procedures.
5. Accessibility	Description of the ACT Aspire accessibility support system, test administration and accessibility levels of support, and accommodations and tools.
6. Test administration	An overview of the processes used to administer ACT Aspire. Users are referred to test administration manuals for more information
7. Test security	An overview of test security and the information security framework governing ACT Aspire.
8. Scores, indicators, and norms	The meaning of scale scores, combined scores, reporting category and writing domain scores, ACT Readiness Levels, progress indicators, development of norms, and norms for scale scores and combined scores.
9. ACT Readiness Benchmarks and Progress Toward Career Readiness	The development and interpretation of college and career readiness indicators, including ACT Readiness Benchmarks and Levels and Progress Toward Career Readiness.
10. Scaling and equating	Construction of the vertical score scales and equating procedures. Note that scaling and mode comparability studies are also documented in the appendix of the technical manual.
11. Reliability and measurement error	Estimates of reliability and standard error of measurement for subject test scores, combined scores, and reporting category scores; rater consistency for writing scores; and classification consistency.
12. Validity evidence	Evidence supporting the validity of proposed interpretations and uses of ACT Aspire scores.
13. Fairness	Four aspects of fairness, with an emphasis on evidence from differential item functioning (DIF) analyses.
14. Growth interpretations	Methodology supporting ACT Aspire's predictions and student growth percentiles; summary data on gain scores and student growth percentiles.

Table A.2. Mean Scores and Percent Meeting Benchmark, by Subject, Grade Level, and Year

Subject	Grade level	Mean score					Benchmark attainment (%)				
		2016	2017	2018	2019	2021	2016	2017	2018	2019	2021
English	3	416.9	417.2	417.2	416.9	415.4	72	73	73	72	62
	4	419.6	419.9	420.4	420.3	419.0	68	70	71	70	63
	5	422.9	422.9	423.1	423.1	421.9	73	72	73	73	69
	6	424.6	425.3	425.3	425.2	423.8	71	74	73	73	68
	7	426.5	427.3	427.7	427.6	426.7	78	79	79	78	76
	8	426.8	427.4	427.9	427.9	426.5	71	73	75	74	70
	9	426.7	427.7	427.9	428.2	427.0	54	58	58	59	55
	10	429.1	430.1	430.2	430.3	428.9	56	60	60	60	55
Mathematics	3	412.9	413.2	413.2	413.3	412.0	56	59	59	62	49
	4	416.0	416.1	415.9	415.9	414.8	54	55	52	54	43
	5	417.8	418.2	417.8	417.5	415.5	48	52	50	49	35
	6	420.8	421.8	420.6	420.0	418.3	55	62	56	54	42
	7	420.4	421.6	421.7	421.6	419.8	44	48	47	48	37
	8	422.8	423.8	424.3	424.5	422.2	38	44	46	48	36
	9	423.2	423.8	424.3	424.5	422.4	28	30	34	36	27
	10	424.8	425.5	426.1	426.2	424.0	22	25	29	29	22
Reading	3	412.4	412.6	412.7	412.9	411.8	35	37	38	38	30
	4	415.0	415.1	415.4	415.6	414.7	41	43	44	45	40
	5	416.9	416.8	417.3	417.6	416.5	34	35	38	41	34
	6	418.9	419.2	419.2	419.1	418.0	46	47	45	43	37
	7	419.7	420.0	420.1	419.9	419.0	36	39	40	39	33
	8	422.4	422.5	422.7	422.9	422.2	48	49	51	53	49
	9	421.5	422.2	421.8	421.6	421.2	36	39	38	37	35
	10	423.1	423.5	423.2	422.9	422.3	35	36	34	32	31
Science	3	414.8	415.1	415.2	415.4	414.0	37	38	39	39	32
	4	417.4	417.8	417.7	417.7	416.6	39	42	42	42	37
	5	419.0	419.7	419.5	419.6	418.3	39	42	41	41	32
	6	421.9	421.7	421.0	421.0	419.5	52	49	46	46	39
	7	421.8	422.5	422.4	422.6	421.6	41	44	43	43	38
	8	423.4	424.2	424.3	424.3	423.3	39	42	42	43	36
	9	424.2	425.0	425.3	425.8	424.9	27	32	33	35	31
	10	426.0	426.5	426.7	427.1	426.1	29	33	33	33	28

Table A.2. (continued)

Subject	Grade level	Mean score					Benchmark attainment (%)				
		2016	2017	2018	2019	2021	2016	2017	2018	2019	2021
Composite	3	414.4	414.7	414.7	414.8	413.4					
	4	417.1	417.4	417.5	417.5	416.4					
	5	419.3	419.5	419.6	419.6	418.2					
	6	421.7	422.1	421.6	421.4	420.0					
	7	422.2	423.0	423.1	423.1	421.9					
	8	424.0	424.6	424.9	425.0	423.7					
	9	424.1	424.9	425.0	425.2	424.0					
	10	425.9	426.6	426.7	426.7	425.5					
ELA	3	417.3	417.6	417.5	417.6	416.1	40	43	42	42	32
	4	419.1	420.1	420.3	420.6	419.6	34	42	43	46	38
	5	421.4	422.3	422.1	422.5	421.0	37	43	41	45	35
	6	423.6	424.5	424.7	424.7	423.4	40	46	47	47	38
	7	423.2	424.2	424.3	424.5	423.4	38	45	45	47	40
	8	424.4	425.5	425.6	425.9	424.8	40	46	48	50	43
	9	424.8	425.5	425.9	425.8	424.9	38	42	44	44	39
	10	426.5	427.1	427.4	427.2	426.2	38	42	44	42	37
STEM	3	414.1	414.4	414.5	414.6	413.3	17	19	20	21	15
	4	416.9	417.2	417.0	417.1	415.9	20	22	21	22	16
	5	418.7	419.2	418.9	418.8	417.2	15	18	16	15	9
	6	421.6	422.0	421.0	420.7	419.2	17	20	15	14	9
	7	421.4	422.3	422.3	422.4	421.0	13	18	19	19	14
	8	423.4	424.2	424.6	424.7	423.0	12	16	17	18	13
	9	424.0	424.7	425.1	425.4	423.9	10	13	14	15	11
	10	425.6	426.3	426.6	426.9	425.3	10	14	15	16	11

Note. The statistics and percentages do not incorporate special scoring rules used by Arkansas for state and federal accountability.

Table A.3. 2019 Readiness Level Transition Percentages, by Subject and Grade Level

Subject	Grade level pair	Prior year = In Need of Support (INS) (%)				Prior year = Close (CL) (%)				Prior year = Ready (RD) (%)				Prior year = Exceeding (EX) (%)			
		INS	CL	RD	EX	INS	CL	RD	EX	INS	CL	RD	EX	INS	CL	RD	EX
English	3–4	44	47	8	1	22	48	25	5	5	28	45	23	0	5	23	72
	4–5	26	60	14	0	11	48	37	4	2	22	56	20	0	4	29	67
	5–6	28	55	16	1	11	48	34	7	2	20	47	31	0	3	20	77
	6–7	25	51	23	2	10	38	42	9	2	13	46	39	0	1	14	85
	7–8	52	32	14	2	28	33	30	9	7	18	38	37	1	2	13	85
	8–9	76	20	3	0	48	35	14	2	16	32	35	16	2	7	22	69
	9–10	73	22	5	0	41	37	18	3	13	29	37	21	1	5	18	76
Mathematics	3–4	47	48	5	0	14	62	24	1	2	30	60	9	0	4	44	52
	4–5	38	58	3	0	15	65	20	0	2	31	60	6	0	5	52	43
	5–6	47	44	8	0	19	53	27	2	2	24	57	16	0	3	33	64
	6–7	72	24	4	0	38	42	17	3	6	28	40	25	0	4	19	77
	7–8	74	22	4	0	39	39	18	4	9	27	38	26	1	5	20	75
	8–9	86	12	2	0	57	29	12	1	22	33	34	11	2	10	30	57
	9–10	88	11	1	0	56	32	11	1	20	37	33	10	2	11	31	55
Reading	3–4	61	30	9	1	15	42	35	9	3	21	46	30	0	4	26	70
	4–5	71	22	6	1	30	39	25	6	7	27	41	25	1	8	29	62
	5–6	69	22	8	2	26	36	26	11	7	23	35	35	1	6	22	70
	6–7	69	25	6	0	26	44	27	3	7	30	50	13	1	10	48	40
	7–8	54	30	15	1	16	32	43	9	4	15	50	31	1	5	31	64
	8–9	77	18	4	1	38	39	18	5	12	30	36	22	2	11	30	57
	9–10	79	17	3	0	42	40	16	2	14	35	39	13	2	13	41	44

Table A.3. (continued)

Subject	Grade level Pair	Prior Year = In Need of Support (INS) (%)				Prior Year = Close (CL) (%)				Prior Year = Ready (RD) (%)				Prior Year = Exceeding (EX) (%)			
		INS	CL	RD	EX	INS	CL	RD	EX	INS	CL	RD	EX	INS	CL	RD	EX
Science	3–4	64	27	9	1	20	41	33	6	7	26	50	17	1	7	33	59
	4–5	65	27	8	1	26	42	28	3	8	27	49	16	1	8	40	51
	5–6	66	23	10	1	26	35	34	6	7	19	49	25	1	4	30	65
	6–7	75	19	5	1	33	35	25	6	9	22	41	28	1	4	21	74
	7–8	77	18	5	0	35	35	25	5	10	23	41	26	1	5	24	70
	8–9	82	15	3	0	45	38	16	2	17	33	37	13	3	9	31	57
	9–10	81	15	3	0	40	40	18	3	13	32	37	17	2	9	27	62
ELA	3–4	76	18	6	1	26	37	28	8	6	21	41	32	0	4	19	76
	4–5	69	24	6	1	20	43	30	7	5	25	44	26	1	5	25	69
	5–6	68	24	7	1	19	40	33	9	4	20	42	34	0	4	20	76
	6–7	72	20	7	1	23	38	30	9	5	20	41	33	1	3	19	77
	7–8	71	21	6	1	21	36	33	10	3	15	41	41	0	2	14	83
	8–9	83	13	3	0	35	37	22	6	8	23	39	30	1	3	16	80
	9–10	84	12	3	0	37	37	21	5	9	25	40	26	1	4	20	75

Note. INS = In need of support; CL= Close; RD = Ready; EX = Exceeding.